

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-1
Perfect score: 15
Sequence: 1 acctcttgatcgac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	6	AR183474	AR183474 Sequence
2	15	100.0	286	9	HSVHL517	273860 H.sapiens r
3	15	100.0	15068	4	MVRYR3	Y07749 M.vision mrn
4	14	93.3	1018	11	CNS06JG2	AL401544 T7 end of
5	14	93.3	1443	6	E15997	E15997 CDNA encodi
6	14	93.3	1626	6	E15998	E15998 CDNA encodi
7	14	93.3	1638	9	HSETBRLP2	Y16280 Homo sapien
8	14	93.3	2411	6	AX018426	AX018426 Sequence
9	14	93.3	10585	1	AE001342	AE001342 Chlamydia
10	14	93.3	16360	4	MAC303116	AJ303116 Tachygllos
11	14	93.3	55739	8	AP000389	AP000389 Arabidops
12	14	93.3	67079	2	AC123621	AC123621 Mus muscu
13	14	93.3	68054	9	AL592300	AL592300 Human DNA
14	14	93.3	69928	2	AC100880	AC100880 Mus muscu
15	14	93.3	73509	2	AL137862	AL137862 Homo sapi
16	14	93.3	83605	2	AC111830	AC111830 Rattus no
17	14	93.3	100406	2	AC122585	AC122585 Rattus no
18	14	93.3	123016	2	AF189001	AF189001 Homo sapi
19	14	93.3	125703	2	AC123338	AC123338 Rattus no
20	14	93.3	160764	9	AC107928	AC107928 Homo sapi
21	14	93.3	161618	2	AC122459	AC122459 Mus muscu
22	14	93.3	168346	2	AC036227	AC036227 Homo sapi
23	14	93.3	172862	2	AL645796	AL645796 Homo sapi
24	14	93.3	176593	2	AC087661	AC087661 Homo sapi
25	14	93.3	179901	2	AC011006	AC011006 Homo sapi
26	14	93.3	181672	9	AC079822	AC079822 Homo sapi
27	14	93.3	187134	2	AC121851	AC121851 Mus muscu
28	14	93.3	189608	9	AL359253	AL359253 Human DNA
29	14	93.3	199650	9	AF252829	AF252829 Homo sapi
30	14	93.3	199810	2	AC091562	AC091562 Homo sapi
31	14	93.3	200355	2	AC019218	AC019218 Homo sapi
32	14	93.3	226715	2	AL596104	AL596104 Mus muscu
33	14	93.3	259854	2	AC098462	AC098462 Rattus no
34	14	93.3	281659	2	AC074159	AC074159 Mus muscu
35	13.4	89.3	39	6	AR116990	AR116990 Sequence
36	13.4	89.3	39	6	AR116991	AR116991 Sequence
37	13.4	89.3	60	6	I68042	I68042 Sequence 11
38	13.4	89.3	217	9	HS52A9R	Z61427 H.sapiens C
39	13.4	89.3	263	9	HS93B11R	Z63943 H.sapiens C
40	13.4	89.3	265	9	HS8F1B01	Z75463 H.sapiens r
41	13.4	89.3	341	5	AF271485	AF271485 Salmo sal
42	13.4	89.3	360	6	I81217	I81217 Sequence 10
43	13.4	89.3	363	9	U00528	U00528 Human immun
44	13.4	89.3	366	6	AX112690	AX112690 Sequence
45	13.4	89.3	372	9	HUMIGHCY	M34678 Human Ig ac

ALIGNMENTS

RESULT 1
AR183474 LOCUS AR183474 15 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 1 from patent US 6342220.
ACCESSION AR183474
VERSION AR183474.1 GI:20227443
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 1 29-JAN-2002;
FEATURES Location/Qualifiers

source 1.15
/organism="unknown"
BASE COUNT 2 a 5 c 4 g 4 t
ORIGIN
Query Match 100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCTCTTGATCGGC 15
|||||
1 ACCTCTTGATCGGC 15
Db 1 ACCTCTTGATCGGC 15
RESULT 2
HSVHL517
LOCUS
DEFINITION H.sapiens rearranged DNA for Ig heavy chain, VH251-D-J, clone
L5-17.
ACCESSION 273860
VERSION 273860.1 GI:1495558
KEYWORDS diversity region; immunoglobulin heavy chain; immunoglobulin
superfamily; joining region; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 286)
AUTHORS Dunn-Walters,D.K.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1996) Dunn-Walters D.K., UMDS St Thomas' Campus,
Dept of Histopathology, Lambeth Palace Road, London SE1 7EH, UK
2 (bases 1 to 286)
Dunn-Walters,D.K. and Spencer,J.
Origins of lamina propria plasma cells
Proc. Ninth Int. Congress for Mucosal Immunol.
JOURNAL
FEATURES
source
1.286
/organism="Homo sapiens"
/isolate="specimen B-2, area 31"
/db_xref="taxon:9606"
/chromosome="14"
/clone="L5-17"
/cell_type="plasma cell"
/tissue_type="ileal lamina propria"
/rearranged
241.273
D_segment 274.286
J_segment 64 a 89 c 69 g 64 t
BASE COUNT 64 a 89 c 69 g 64 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 286;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 ACCTCTTGATCGGC 15
|||||
37 ACCTCTTGATCGGC 51
Db 37 ACCTCTTGATCGGC 51
RESULT 3
MVR3R3/c
LOCUS
DEFINITION M.vision mRNA for ryanodine receptor 3.
ACCESSION Y07749
VERSION Y07749.1 GI:1526614
KEYWORDS ryanodine receptor 3; RYR3 gene.
SOURCE Mustela vision.
ORGANISM Mustela vison
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
Mustela.
REFERENCE 1 (bases 1 to 15068)

AUTHORS Marziali,G., Rossi,D., Giannini,G., Charlesworth,A. and
Sorrentino,V.
TITLE CDNA cloning reveals a tissue specific expression of alternatively
spliced transcripts of the ryanodine receptor type 3 (RYR3) calcium
release channel
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 15068)
AUTHORS Sorrentino,V.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1996) V. Sorrentino, DIBIT, Istituto Scientifico
San Raffaele, via Olgettina 58, I-20132, Milano, ITALY
FEATURES
source
1.15068
/organism="Mustela vison"
/db_xref="taxon:9667"
/cell_line="CCL64"
1.14580
/gene="RYR3"
1.14580
/gene="RYR3"
/codon_start=1
/product="ryanodine receptor type 3"
/protein_id="CAA69029.1"
/db_xref="GI:1526615"
/db_xref="SPTREMBL:Q95201"
/translation="MAEGEGEGEDEIQFLRTEDVEVYLQCIANIHKQKFCIAEGIG
NRLCFLEPTSEAKVPPDLCVNFVLEQSLSVRALQELMANTGEGGGAQGGHRT
LLYGHAILLRHSFGSMYLLCLTTSRSQTDKLADYGLRENATGEGACWTIHPASKORS
EGEKVRIQDDLILSVSSERYLHLVSNGNVQDASFMQTLWNVHPTCSGSSIEGYL
LGGHVRLEFHGDECLTIPSPDQNDQAQHRRIEYAGAGATRARSLMRDPLRISWSGS
NIRWQAFRLRHITTTGHTLALTEDQGLLQDRAKSDTSTAFSFRSKEIKEKLDSSH
KRDIENGVPKIKYGDVSVCAQDAGLWVTVKADAKTSRILGKPKKRVILHOGHMD
DGLTLQRCQREESQAARIIRNTALFSQFVSGNNRAAPVTLPIQEVLTQLQDLAYE
OPPEEOMHEDKONKLRSLKRNQLKEEGMALVINCIDRLNINVSVAHFAGIAREE
SGMAKKEILNLKLLAALIRGNRNCAQFSNLDMLISKLDRLESSGILLEVLCIL
IESPEALNLAEGHIKSIIELDKHGRNHKVLVLCISLCINGVAVRANONLIDNLL
PRNLILQTRLINDVTSIRBNILGVAEGSAQKKWFELIIDQVDEPLTAEPYHLRV
GWASSSGYAPYPGGEGGNGVGDLLYSYGFGLHLSGRIPRAVASINLHLRSDD
VVSCLDLGVPISIFRINGQPVQGMFENFNTDGLFFPVVSFAGVYKVFLLTGRRHGEF
KFLPPSGYAPCYEALIPKEMRLPEPKYKRWDAEIRDLVLTQSLSQASFIPCIDT
SOVLPPHLEKIRDLRAENIHELMGNKIELGWTGKIRDNKRQHPCLVEFSKLPET
EKNYNLQMSSTETLKTLLALGCHIAHVNPAEEDLKVKLPKNYMSNGYKPPALDLS
VKLPPQEVLDKLAENAHNWAQDKRIQGVYGIQDQDLKKNRNPRLVPYALLDERTK
KSNRDSLRBAVRTFVGYGYNTEPSQDOELAPAVEKVSIDKIRFRVEMSVAVRSKMY
FEFEVVTGDMRVGWARPGRPDIELGADDQAFVEGSRGQRWQSGYFGRTWQGD
VVGCMINLDDASMIETLNGELLITNKSEIAFADYEIENGFPICSLISLOIGRMNLG
MDASTEKFYTMCGLOEGFEPPFAVMNMNDVAMVFSKLPTEFVNVPKDPHPIEVRIQEI
LSHTTQCYATRIFFAGQDSCVWVWVTPDYHLSEKFDLKNKCTVTLGDERGRV
HESVKNRNCYMWGCDIVATSQSSRSNVDLEICLVLDLAMGMLSFANGRELGTQYQ
IOPNTKVEPFAVFLQPTSTALFQELGKIKNAMPLSAIFKSEENPVPQCPRIDVOT
QVPLVMSRMPSSFLKVEVEGVSERHGWAVQCLEPIQOMALHPIENRCVDILELCEQE
DLMQFHYHTLRLYSAVCALGNSRVAYALCGHVDLSQLEYAIDNKYLPGLLRSGFYDL
ISIHLANAKERKLMKNETIIPITSATRKIRLYPDESKRHGLPGVGLRTCLKGRFES
TPCFVVTGEEHOKOSPEIPELILKTALSMLTEAVQHSGAHIRDPVGGSVFEQFVPL
KLIGTLVMGVFDDDDIRQITLILIDPSVFGESHVETEDGAKEKEVYQVEEKAVEAGEK
AGKEADYKGLQTRLPESVKLMQCELSYLDCPELOHREVAIVAFCGDIYVSKIQANOK
SRYNELMALMSAALTARKTRFRSPQEQIMNLLNFQLGENCEPCPEEIREELYDFH
LISQTMIRWAQEDQIODAELVRMFSLLQRYDSIGELLQALRKTYTISQASVSDTIN
LIAALGOIRCLISVRMGKEEELIMINGIDIMNNKVFYOHPNLMRYLGMHETVEMV
DVILGAESQIAFPKMWASCCRFICYCRISRONKAMFELSTYLLNESSVGLASPSMR
GSTPLDVAASSMDNNEILAGIEPDLEKVVYIAGCGIQCSPMLAKGYPDVGNPPI
EGERYLSFLRAVFVNSEVENEDGEDEIVHMGNAIMSFYALIDILGRCAPEMHLIOT
GKGEAIRIRSLIRSLVPTEDLVGIIISIPKLPISLUNKDGSISEPDMAANFCPDHKAPMV
LFLDRYGIKQDTFILHLLEVGLDPLIRASASLDIVSLSTTEAALNLRILNLSAVLPL
LITRCAPLFAGTEHYTSLIDSTLQITRYRLSKGRSLTKAQRDTIEECLAICNHLRPSML
QQLRLRVFDVQPLNENYCKMPLKLTITNHYEQCKKYCLPSGSGSLAVEEELHLETEK
LFWGLTDSLHKKYDPLDFRMALPCLSAIAGALPPDYLDTRISATLEKQISVDADGNE
DPKPINNMNFSLEPKWEYIYTKYAEHSHDKWADKSONGKYGISLDENYKTHPLIRP
FKTLTEKEKEIYRWPARESLKTMLAVGWSVERTKEGALVQORENKLRSVSQASQGN
SYSPAPLDSLNVLSRELQGWVEVAENHNIMAKKKKLELESKGGSHPLLVPIYDTL

TAKEKFRDREKAQDLKFLQVNGIVSRGVKDMELDASSMEKRFAYKELKILKYVDS
AQEFIAHLEAVISSGKTEKSPHDOEIKFEAKVILPLVDQFTNHILYFSSPLKPLSS
SGYASHKEKEMVASLFCKLALVHRISLFGSDSTTWVSLHILAQTLDTRTYMKS
ELVKAGLRAFENAEDELEKTESNKLKGFTHSFTQIKGVSONINTVALPLPISI
FEHVAQHOFVLDLLGDVQISCIYHLCISLSTGKNIIYERORPALGECIASLAAI
PVAFLPTLNRYNPLSVENKTPRERSILGMPDVEEMCPDIPOLEGLMKETSDLAES
GARVTEMPHVIEVILPMLCNLYSYWMERGPEHLPESTGPMCTKVTSHELSVILGNILK
IINNNGIDEASMMKRIAYVAQPIISKARPDILKSHFTPTLEKIKKAAKVAQVEEOL
KADTKGDTQEAELLILDEFANVLCRDLYAFYPMILIRYVDNNRSMWSKSPADSQLRG
VAEVEFILMCKSHNFKREONFVIONENINLAFLTGDSKSKMSKSGQDQERKTKRR
DLYSIOTSLIVAALKMKPLIGLNMCTPGDOELISLAKSRYSYRDTDEEVKEHLNNLIH
LOEKSDPAVKWOLNLYKDVLSKEEFPNAEKTERVORISAAVHLEQVEQPLRSKKA
VMHKLISKORKRAVAVACFRMAPLYNLPKHINNFEISTQFVWLEKNEKTOYDRLLP
ILMKSPKVEEEEEEEMEKQDPLHQIILHFSRNALTERSKLEDDPLYTSYSMAKASC
OSGEDEEEEDKEKTEFEKEMEKOKTLYQARLHERGAEMVLOMISASKGEMSPMVV
ETLKLGIALLNGNAGVQOKMLDYLEKKDAGFEQSLGLMOSCVLDLNAFERONKA
EGLGNVTEEGTREKVLONDEFTRLDFRLQLLCEGNSDFQNELRTOMGNTTINVI
STVDYLLRLQESISDFYWYSGKDIIDESQHNEFKALAVTKOIFNSLLEYIQGPCIG
NQOSLAHSRLMDAVVGFHLHFANMQMLSDSSQIELLKLDDLLQDMVVMLSLLEG
NNVNGTIGKOMVDTLVESSTVNMILKFDMLKLKLDSSDTEKEYDPDGKGLTSKK
EFQKAMEGOKOTQSEIDFLISCAEADENDMFNVDFRHEPAKIDGFNAVALTN
LSEHMENDSRILKCLDPAGSVLVNFEYPIRIEMGAKTERVYFEISSERTQOMEK
POVKESKROFIFDVVNEGGEQEMLEFVFCEDTIFEMOLASQISETDSAERPDEEE
EDEDSSVYLEIEGEEDDKSPESASAFAMACASVRNANFLKATLKNLRQYRYVK
KMTVKELVKVFFSFFMMLFVGLFRLFTILGIFQILNMNVFEGGLVEGAKNIRVTKI
LGDMPDETOEGIHDDALDAERAETEPGPPELILHFVKERGDADIMSDLFGFHPKKE
GGLKHGPEGGLDGLSEIIGKDEPTLESTVRKKRKAQAEMKAAHEAGKVEPEKTDL
EDGEKEDTAKEEOAEALMWADVKKRRGQKVEPEAFMANEFGLEIYQTKLHY
LARNEYNLRFLLFAFAINFILFYKVTBEEVEEETEDVANLWNSFTDEEEEAAMF
FVLOESTGYMAPALRALAIHVTIVSLVCVGYCLKVPVLFKREKEIARKLEFDGLY
ITEQPSDEDDIKQWDRLVINTPSPNNYWDKFKRVKINKYGDLLYCAERILLEGLDK
NALDFSPVEASKAEAASIVNSLSDIMKYHMKLGVTFTNSFLYLAWYTTMSVLGHY
NNEFFFAHLLDIAMGFKTLRTILSSVTHNGKQLVLTGLLAVVYLYTVAAFNFRKF
YNKSEDDDERDMKDDMTCYLFHMVYGVRAAGGIGDEIEDPAGDPYEMYRIVFDTITF
FFEVIVILLAIIOGLIIDAFGLRQOEQVREDMETKCFICIGINDYFDTTPHGFETH
TLQEHNLANVLFMLYLINKDETEHTGOESYVWKMYQERCMDFEPAGDCFRKQYEDQL
G"

BASE COUNT 3969 a 3667 c 3990 g 3442 t
ORIGIN

Query Match 100.0%; Score 15; DB 4; Length 15068;
Best Local Similarity 100.0%; Pred. No. 3.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
Db 1316 ACCTCTTGATCGGC 1302

RESULT 4
CNS06JG2 1018 bp DNA linear STS 10-JAN-2001
LOCUS T7 end of clone AS0AA028E04 of library AS0AA from strain CLIB 533
DEFINITION of Saccharomyces bayanus, sequence tagged site.
ACCESSION AL401544
VERSION AL401544.1 GI:12159319
KEYWORDS STS.
SOURCE Saccharomyces bayanus.
ORGANISM Saccharomyces bayanus.
REFERENCE Saccharomycetales; Saccharomycetaceae; Saccharomyces.
AUTHORS 1 (bases 1 to 1018)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neugeglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
JOURNAL yeast species for molecular evolution studies
MEDLINE FEBS Lett. 487 (1), 3-12 (2000)
PUBMED 20584711
11152876

REFERENCE 2 (bases 1 to 1018)
AUTHORS Bon,E., Neugeglise,C., Casaregola,S., Artiguenave,F., Wincker,P.,
Aigle,M. and Durrens,P.
TITLE Genomic exploration of the hemiascomycetous yeasts: 5.
JOURNAL Saccharomyces bayanus var. uvarum
MEDLINE FEBS Lett. 487 (1), 37-41 (2000)
PUBMED 20584715
11152880

REFERENCE 3 (bases 1 to 1018)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This STS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
source location/Qualifiers
1. 1018
/organism="Saccharomyces bayanus"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA028E04"
/clone_lib="AS0AA"
/note="end : 17"

misc_feature complement(<18..>422)
/note="similar to Saccharomyces cerevisiae ORF YLL043w [FPS1 : glycerol channel protein]"
/evidence=not_experimental

BASE COUNT 232 a 213 c 243 g 328 t 2 others
ORIGIN

Query Match 93.3%; Score 14; DB 11; Length 1018;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGG 14
Db 109 ACCTCTTGATCGG 122

RESULT 5
E15997/c 1443 bp DNA linear PAT 28-JUL-1999
LOCUS cDNA encoding human G protein-coupling receptor protein.
DEFINITION E15997
ACCESSION E15997
VERSION E15997.1 GI:5710680
KEYWORDS JP 1998127289-A/1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 1443)
Hinumata,K., Fukuzumi,M. and Kawamata,Y.
TITLE NEW G PROTEIN COUUGATE TYPE RECEPTOR PROTEIN AND ITS DNA
JOURNAL Patent: JP 1998127289-A 1 19-MAY-1998;
TAKEDA CHEM IND LTD
OS Homo sapiens (human)
PN JP 1998127289-A/1
PD 19-MAY-1998
PF 29-OCT-1996 JP 1996286823
PI HINDMA KUNIJU, FUKUZUMI MASASHI, KAWAMATA YUJI PC
C12N15/09,C07H21/04,C07K14/705,C07K16/28,C12N1/21,C12P21/02, PC
C12P21/08,
PC C12Q1/68,G01N33/53,G01N33/566//A61K38/00,A61K39/395,

[illegible]

BASE COUNT	272 a	536 c	488 g	330 t	/db_xref="taxon:9606"
ORIGIN					
Query Match	93.3%;	Score 14;	DB 6;	Length 1626;	
Best Local Similarity	100.0%;	Pred. No. 1.7e+03;			
Matches 14; Conservative	0;	Mismatches	0;	Indels	0; Gaps 0;
QY	2	CCTCTTGATCGGC	15		
Db	321	CCTCTTGATCGGC	308		
RESULT 7					
HSETBRLP2/c					
LOCUS	HSETBRLP2	1638 bp	mRNA	linear	PRI 15-APR-1998
DEFINITION	Homo sapiens mRNA for G protein-coupled receptor ETBR-LP-2.				
ACCESSION	Y16280				
VERSION	Y16280.1 GI:3059117				
KEYWORDS	ETBR-LP-2 protein; G protein-coupled receptor.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1638) Valdenaire,O., Giller,T., Breu,V., Ardatti,A., Schweizer,A. and Richards,J.G.				
TITLE	A new family of orphan G protein-coupled receptors predominantly expressed in the brain				
JOURNAL	FEBS Lett. 424 (3), 193-196 (1998)				
MEDLINE	98198552				
PUBMED	9539149				
REFERENCE	2 (bases 1 to 1638) Valdenaire,O. Direct Submission Submitted (22-JAN-1998) O. Valdenaire, F. Hoffmann-La Roche Ltd., Prpv 69/348, Grenzacherstrasse, 124, CH-4070 Basel, SWITZERLAND				
AUTHORS	location/Qualifiers				
TITLE	1. 1638				
JOURNAL	/organism="Homo sapiens"				
MEDLINE	/db_xref="taxon:9606"				
PUBMED	/clone="lambda GT10"				
REFERENCE	52. .1497				
AUTHORS	/function="G protein-coupled receptor"				
TITLE	/codon_start=1				
JOURNAL	/product="ETBR-LP-2 protein"				
MEDLINE	/protein_id="CAA76153.1"				
PUBMED	/db_xref="GI:3059118"				
REFERENCE	/db_xref="SPTREMBL:O60883"				
AUTHORS	/translation="MRWLPLAVSLAVTAVGLSRVSGAPLHLGRHRAETQEOQSR KRGTDEEAKGVQOYVPEEMAEYPRPIHPAGLOPTKPLVATSPNPKDGTDPDSGQEL RGNLTGAPGRLQIONPLYPVTESYSAYAIMLALVFAVGIVGNLSMCTVMHSY LKSAMNSILASIALMDLVLFFCLPIVIFNETTKORLLGDVSCRAVPMEVSSLGVT FSLICALGIDREHVAATSLPKVPIERCOSILAKLAVIVGSMILAVPELMLQAOEP APTMTGLDSCIMKPSASLPESLSLVMTYONARMMWYFGCYFCLPILETVCQLVTR VRGPPGRKSECRASKHEOCESQNSTVVGILTVVAFCTLPENVCNIVAVISTELTRO TLDLGLINQSTFFKGAITPVLLCLICRPLGQAFLDCCCCCECGGASASAANG SDNKLKTEVSSSYFHKPRESPLPLPLGTFC"				
BASE COUNT	273 a	526 c	493 g	346 t	
ORIGIN					
Query Match	93.3%;	Score 14;	DB 9;	Length 1638;	
Best Local Similarity	100.0%;	Pred. No. 1.7e+03;			
Matches 14; Conservative	0;	Mismatches	0;	Indels	0; Gaps 0;
QY	2	CCTCTTGATCGGC	15		
Db	189	CCTCTTGATCGGC	176		
RESULT 8					
AX018426/c					
LOCUS	AX018426	2411 bp	DNA	linear	PAT 07-SEP-2000

DEFINITION Sequence 1 from Patent EP0943685.
ACCESSION AX018426
VERSION AX018426.1 GI:10042588
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 2411)
TITLE Kroeger, B.D. and Otterbach, B.D.
JOURNAL G-protein coupled receptor from human brain
Patent: EP 0943685-A 1 22-SEP-1999;
BASF AG (DE)
FEATURES
source location/Qualifiers
1..2411
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..19
20..1465
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC07699.1"
/db_xref="GI:10042589"
/translation="MRWLMPPLAVSLAVILAVGLSRVSGGAPLHLGRHRAETQEQSRS
KRGTDEEAKGVQOQVPEEMAEYRPRIHPAGLQPTKPLVATSPNPKDGGTSDGOEL
RGNLTGAPGQRLQIONPLYPTESSYSAIAMLALVFAVGIVGNLSVCIWHSY
LKSAMNSILASLAWDFLVLFECPLIVFENETKORLLGDVSCRAVPMEVSSLGVT
FSLCALGIDRFHVATSTLPKVRPIERCQISILAKLAVIWSMTLAVPELLMQLAQEP
APTMTLDSICMKPSASLESLSLVMTYQNAWMWYFGCYFCLPILFTVQCQLVTR
VRPPGRKSECRASKHEQCESQNSTVGLTVVYAFCTLPENVCNIVAVYLSLELTRQ
TLDLLGLINQSTFFKGAITPVLILCICRPLGQAFIDCCCCCECGGASASAAANG
SDMKLTKEVSSSIYFHKRPSPLPLGLTPC"
3'UTR 1463..2411
BASE COUNT 397 a 775 c 659 g 580 t
ORIGIN
Query Match 93.3%; Score 14; DB 6; Length 2411;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2 CCTCTTGATCGGC 15
|||||
Db 157 CCTCTTGATCGGC 144
RESULT 9
AE001342 10585 bp DNA linear BCT 30-OCT-2000
LOCUS Chlamydia trachomatis section 69 of 87 of the complete genome.
DEFINITION AE001342 AE001273
ACCESSION AE001342.1 GI:3329166
VERSION
KEYWORDS
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS 1 (bases 1 to 10585)
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V. and Davis, R.W.
TITLE Genome sequence of an obligate intracellular pathogen of humans:
JOURNAL Chlamydia trachomatis
MEDLINE Science 282 (5389), 754-759 (1998)
PUBMED 99000809
9784136
REFERENCE 2 (bases 1 to 10585)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 3 (bases 1 to 10585)
AUTHORS Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,

Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES
source location/Qualifiers
1..10585
/organism="Chlamydia trachomatis"
/strain="D/UW-3/CX"
/db_xref="taxon:813"
88..2391
/gene="CT711"
88..2391
/gene="CT711"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68306.1"
/db_xref="GI:3329167"
/translation="MSIQPTSLTKNITTAALAGEQVDAAYVMPQAVFFPQOLDEKS
KGLQALGLEVDLEKFIPLSEKSPPTTGTTSKISADGIEIVGELSSETILADPN
KAAQVFGELADSEDDWLRLESENGIOPPTAIEEIVKQTELTNLRNKLKQSLT
DDEYTKLYAIPQNFVKEIESLKNENVRILPKSKVTNFWQNTMLTNSVTSLEPVD
AMNTMAEVSLEYIERATEAAKLIREITNTIKDIFNPVMDVREQTGIFGLGAEYNAL
GNMIQSLSEAGLFQRLMSRTAYDEIGALYPKNDKNEDVIHTAIDYVNSLADKAN
EQVKLNGLSLVYAYASTLGFAKKQVENNAQASFTDYTNELNQEIQYWTPRETSFN
ISNOALQTFEKNPSADYNGVYLFDNKGLETNLFNPTFEFDVYSLMTADPTKMSRDY
NKVITASSESLQKINQAITAMELALAECCGAKKALEPSSLNENAMVEAKKTVEETSP
IOMVYSLMDKYLPNQOYILETIGSQMTFSNKAARYLNDILAYAVSFQADAVYSLG
MYLRQMNQDEPPEVISRANDTVKKEIDRSRADLFHCKAIEIKELVTSVNADTELTS
SORAELETFETLASYAFEEFENLYHNLNRYVMVSKVQISGVSKDEVDDEAFTAKIGSKEF
DTWIOQLTFRESAVIEGGRNGVMPGGEQVLOSLESKQODYTSFNQNLALQMESAA
IQQEWTVMAALALMNQIFAKLIRRFK"
2418..3590
/gene="CT712"
2418..3590
/gene="CT712"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68307.1"
/db_xref="GI:3329168"
/translation="MRNHPPEGQKYFVPPIETAPRRERVSPEAVYADYTQLHEAAT
YLQVFODLNDADHQLGKNEFVESLRQDFLKTGSEMSLMQALWTEESQREARKRERKE
LQQLLESKVLGPQALTTAEELHPVDDSIYNKMPQSAFAYILLDKYIPAQEBALYALA
RELSESGYAOFTLFSVPLELVKSFNNAPRYVNLGSYIGOTAGTANFKYGYOMVLDREYET
ETGOLRKDIKNAENAKQOLAQIINKINEANNSLTTEHKTQLKDMANGYIOTLDVCISQM
QELSTGLRGLSFIPIGRDEXSPAYEIMGSSFSVTVLQNLGKVVDEGINISSGETKGL
LNFFTYFLADVQNFQDLAQTNQMLMELQMRAMHQQWSLVYASLKLHNVYRTLAAS"
complement(3616..4638)
/gene="porB"
/note="CT713"
complement(3616..4638)
/gene="porB"
/codon_start=1
/transl_table=11
/product="Outer Membrane Protein Analog"
/protein_id="AAC68308.1"
/db_xref="GI:3329169"
/translation="MSSKLVNYLRLTFLSFLGIASTSLDAMPAGNPAFPVPIGINIEQ
KNACSFDLNSYDVSLSLGNLKLFCFGDYIFSEEAQVKDVPVTSVTAGVGSPSDI
TSTTKTRNFDLVNCLNLTNCVAVAFSLPDRSLSAIPLFVDSFEVKVGLKQYRLPMN
AYRDTSEPLNSESEYTDGMIEVQSNYGFVMDVSLKKVIWKDGVSEFGVADYRHASC
PIDYIANSQANPEVFIADSDGKLNFKEWSYCVGLTFTYVNDYVLPYLAFSIGSVSRQA
PDDSFKKLEDRFTNLKFKVRKITSSHGNICIGATNYVADNFYINVEGRMGSQRAVNV
SGGFQF"
complement(4771..5775)
/gene="gpda"
/note="CT714"
complement(4771..5775)
/gene="gpda"
/codon_start=1
/transl_table=11
CDS
gene
CDS
gene
CDS

/product="Glycerol-3-P Dehydrogenase"
/protein_id="AAC68309.1"
/db_xref="GI:3329170"
/translation="MKETIAYLGWGMGFSLANLLANNCHRVVGMARNPALIEQLSVQ
RRHPAPHISIPQNLSEFTSHMEALDGMTIVEGVSAGMRPVLTQLKALTELRLVPLV
ITSKGIQONTGLLSEIALEIFGRPAOHLGYLSGPSIAEVLGCCPSVVISAYNP
TLKQIHRAFLTEFRVYNSDLGALKNVIAICGISDGRFEGDNASKGLVTR
GLHEIRKFAITMGCRPDTLNLGAGLDCTTSFSAFSRNTLFGKLLAEGLTPEQAKTK
IGMVEGVYTAASHQIATHRIDMPTTTSVRYLVYENLDIQEIAQLLRDTKEEYL
"
complement(5772. .7139)
/gene="CT715"
complement(5772. .7139)
/gene="CT715"
/codon_start=1
/translation="agx-1 Homolog-UDP-Glucose Pyrophosphorylase"
/product="agx-1 Homolog-UDP-Glucose Pyrophosphorylase"
/protein_id="AAC68310.1"
/db_xref="GI:3329171"
/translation="MTDSSYVDLSFLDQLLPIDQEHLLLEWPSLSPQORLRIGTQIA
QIDIPFLRQOALLQNPQASHQETPLSPVHYAGDNPAYAQLGFLQKVGCVVLA
GGGSRLLKFDGPKGLYPVSSVKRPYQOLVAEKVAASKRWGRPLAINTPLNHKQ
TLSTFATNDYFNLSPSQVDFPCQPLWLLSLGDLFLESEDRSLGPTGNGCLSTLLQ
SSGIWDKMHQAGIEMVSVIPIDNPALPDRLELVGFHAAEHNDVTIKTTLROSAQEDV
GVLIELAKQKTAVEYSTLTTERCKVKTTEGDLTYKLANIGYCLSMDFLAQTAYQPL
PLYKANKHAKQHPSTTEKNWKFEEFTFDLFOYSEHSAIVYPRHCEVAPLKNVEGN
HSPATVREAMRKREHALFTAVTERKLSPTNIFLEADFYPPSSHTSLEWETKIFQET
IEAS"
complement(7151. .7516)
/gene="CT716"
complement(7151. .7516)
/gene="CT716"
/codon_start=1
/translation="hypothetical protein"
/product="hypothetical protein"
/protein_id="AAC68311.1"
/db_xref="GI:3329172"
/translation="MNKLIQDLKSLTLEFKKRTLETVKKALSTIEHRLQIQEHI
AKISLTRHKQFLCRSYTHEYDQHLHLEQREQTSLYKHQTLKTSKDAYGDIQKQDQ
RKIEKIHDSKYPKISANN"
complement(7509. .8813)
/gene="f11f"
/note="CT717"
complement(7509. .8813)
/gene="f11f"
/codon_start=1
/translation="flagellum-specific ATP Synthase"
/product="flagellum-specific ATP Synthase"
/protein_id="AAC68312.1"
/db_xref="GI:3329173"
/translation="MTHLOEETLLIHOWRPYREGILSRIGSLIEAQLSACLGELC
QISLSRSDPILAEVIGIHNTLLLATPIYLLAIGAEVPLRKPASILPSNHLGRV
LDGFNPIDGPDLPKTNLSPLSSPSPMSRTPIOEVPTGIRADALLITIGQORV
GIFSEPGGKSSLSITIAKGSQQTINVALIGERGREVDYVQKREGLAQRTVITA
STAYETAASKVIAGRAAITIAEYFRDQGARVLTMDSLSRWIESLQEVATARGETLST
HHYASVFHHVAEFLERAGNNDKSTSFYALHYNHPDIFTDYKSLLDGHFFLSP
QEKSESSPINVLTSLSSRSQALPHEHYAAQELSLKAYHEDIDYIQGAYVSGQ
DAHLDRAIRLPSVKQFLSQPYSHYSATIHETIEQLCQLLKEH"
complement(8887. .9411)
/gene="CT718"
complement(8887. .9411)
/gene="CT718"
/codon_start=1
/translation="hypothetical protein"
/product="hypothetical protein"
/protein_id="AAC68313.1"
/db_xref="GI:3329174"
/translation="MEDYVASPHLRSLACIDNPQPIETPLFEQALSHLISLQIVF
RKLNVHLSEIEKLSOKLPELLELAVLCEKFLYKRLACTEELALLISALQHHLAT
YAVSPIKIGLHPEDLSNLSKWLILHDVPLKNIETIADPLCKKASVKIELPSGILROD
IGBELSHLSVLTP"
complement(9416. .10420)
/gene="f11f"

CDS
/note="CT719"
complement(9416. .10420)
/gene="f11f"
/codon_start=1
/translation="flagellar M-Ring Protein"
/product="flagellar M-Ring Protein"
/protein_id="AAC68314.1"
/db_xref="GI:3329175"
/translation="MPVLPFLKKNIAVTKSLGYLLAAILIGFTIMLYKPPSPQPTPV
ASTEKPPSHWLKLSHLGNLQISIEIQAKKEQLEKDLTFEPVLAQVALSQEEDSLAEI
SVILSLPOASTLSPSLVHSITDYLTFRVPGLTKEHITLSDQHGNDYSPLEQSNLTLT
TSLERSLQTLTQTHFALNYIPVADEGHLLQDLYDELYLNTLPKARVKLLSHMOEILS
AFPEMPSVDIVPEFLKPVAKTSRLSSIVSLITIVLSTGLIGFATFYLAFTYDHSV
QKEKIQISINIPKLIEMMKRESPEKVALILSLDSAKAEELLNKLPEEMKSAVLKLR
T"
BASE COUNT 3114 a 1914 c 2513 g 3044 t
ORIGIN
Query Match 93.3%; Score 14; DB 1; Length 10585;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCCTCTTGATCGG 14
Db 8395 ACCCTCTTGATCGG 8408
RESULT 10
MAC303116 16360 bp DNA circular MAM 23-JAN-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
MAC303116 16360 bp DNA circular MAM 23-JAN-2002
Tachygylossus aculeatus complete mitochondrial genome.
AJ303116
AJ303116.1 GI:17932764
12S ribosomal RNA; 12S rRNA gene; 16S ribosomal RNA; 16S rRNA gene;
ATP synthase subunit 6; ATP synthase subunit 8; atp6 gene; atp8
gene; col gene; co2 gene; cox3 gene; cytb gene; cytochrome b;
cytochrome c oxidase subunit 1; cytochrome c oxidase subunit 2;
cytochrome c oxidase subunit 3; internal transcribed spacer 1;
ITS1; NADH dehydrogenase subunit 1; NADH dehydrogenase subunit 2;
NADH dehydrogenase subunit 3; NADH dehydrogenase subunit 4; NADH
dehydrogenase subunit 4L; NADH dehydrogenase subunit 5; NADH
dehydrogenase subunit 6; NADH1 gene; NADH2 gene; NADH3 gene; NADH4
gene; NADH4L gene; NADH5 gene; NADH6 gene; transfer RNA-Ala;
transfer RNA-Arg; transfer RNA-Asn; transfer RNA-Asp; transfer
RNA-Cys; transfer RNA-Gln; transfer RNA-Glu; transfer RNA-Gly;
transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu(CUN); transfer
RNA-Leu(UUR); transfer RNA-Lys; transfer RNA-Met; transfer RNA-Phe;
transfer RNA-Pro; transfer RNA-Ser(AGY); transfer RNA-Ser(UCN);
transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer
RNA-Val; transfer RNA-Ala gene; transfer RNA-Arg gene; transfer RNA-Asn gene; transfer RNA-Asp
gene; transfer RNA-Cys gene; transfer RNA-Gln gene; transfer RNA-Glu gene; transfer RNA-Gly gene;
transfer RNA-His gene; transfer RNA-Ile gene; transfer RNA-Leu(CUN) gene; transfer RNA-Leu(UUR)
gene; transfer RNA-Lys gene; transfer RNA-Met gene; transfer RNA-Phe gene; transfer RNA-Pro gene;
transfer RNA-Ser(AGY) gene; transfer RNA-Ser(UCN) gene; transfer RNA-Thr gene; transfer RNA-Tyr
gene; transfer RNA-Tyr gene; transfer RNA-Val gene.
SOURCE
ORGANISM
Tachygylossus aculeatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Monotremata; Tachygylossidae; Tachygylossus.
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21592585
11734900
2 (bases 1 to 16360)
Janke, A.
Direct Submission
Submitted (21-DEC-2000) Janke A., Dept. of Genetics, Lund
University, Soelvegatan 29, S-22362 Lund, SWEDEN
FEATURES
Location/Qualifiers

source 1. .16360
/organism="Tachyglossus aculeatus"
/organelle="mitochondrion"
/db_xref="taxon:9261"
/tissue_type="liver"
/dev_stage="adult"
/country="Australia:Melbourne, Healesville Sanctuary"
gene 1. .72
/gene="tRNA-Phe"
tRNA 1. .72
/gene="tRNA-Phe"
/product="tRNA-Phe"
/note="codon recognized: UUC"
/anticodon=(pos:35. .37,aa:Phe)
gene 73. .1022
/gene="12S rRNA"
rRNA 73. .1022
/gene="12S rRNA"
/product="12S ribosomal RNA"
gene 1080. .2646
/gene="16S rRNA"
rRNA 1080. .2646
/gene="16S rRNA"
/product="16S ribosomal RNA"
gene 2648. .2723
/gene="tRNA-Val"
/note="tRNA-Leu (UUR)"
tRNA 2648. .2723
/gene="tRNA-Val"
/product="tRNA-Val"
/note="codon recognized: UUA"
/anticodon=(pos:2684. .2686,aa:Val)
misc_feature 2724. .2807
/note="H-stand conserved spacer sequence, CSS"
/function="unknown"
gene 2808. .3764
/gene="nadh1"
CDS 2808. .3764
/gene="nadh1"
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 1"
/protein_id="CAC88010.1"
/db_xref="GI:17932765"
/translation="MFLINLLLLIVPVLAVAFLLIERKILGYMOFRKGPNIIVCPHG
LQPIADAVKLFKEPLRLTSSIYMFILAPILSLALTIVPLPMLPIDNLGL
LFLISVSGLSVSIILMSGMNSKVALTGALRAVAQTISYEVTFAIILSIMLINGSE
TLTPTNLTOEFMWLVVPTWDLMLMFISTLAETNRAFDLTEGESELVSGFVVEYAAG
PRAMFFLAEYANIIIMNALVILFEGTYHLIFLELSTTNFMVKITMLTSLFLWVRAS
VPRFRYDQLMHLWKNFLLPTLVTCIWMYIMFPTMLSGTPPQM"
gene 3765. .3833
/gene="tRNA-Ile"
tRNA 3765. .3833
/gene="tRNA-Ile"
/product="tRNA-Ile"
/note="codon recognized: AUC"
/anticodon=(pos:3794. .3796,aa:Ile)
gene complement(3831. .3902)
/gene="tRNA-Gln"
tRNA complement(3831. .3902)
/gene="tRNA-Gln"
/product="tRNA-Gln"
/note="codon recognized: CAA"
/anticodon=(pos:3867. .3869,aa:Gln)
gene 3902. .3970
/gene="tRNA-Met"
tRNA 3902. .3970
/gene="tRNA-Met"
/product="tRNA-Met"
/note="codon recognized: AUA"
/anticodon=(pos:3923. .3925,aa:Met)
gene 3971. .5012
/gene="nadh2"

CDS 3971. .5012
/gene="nadh2"
/note="TAA stop codon is completed by the addition of 3' A
residues to the mRNA"
/codon_start=1
/transl_except=(pos:5012,aa:TERM)
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="CAC88011.1"
/db_xref="GI:17932766"
/translation="MTPMTNLIIMSSILMGTMITLTSSHWLLMWMGLEINTLAIPPL
TSKKPRSTESAIKYFLTQATASMLMFASFASLNTWLTGHWTLMOINNVPSSIMITFAL
AMKGLAPHYWVPEVLQSGSPILSGMILLTWOKLAPISIIYQISPTLNMNTLLILAIS
SILGGWNGLNOTQLRKIMAYSSIAHGMWMIIVIIYFPLTILNLVLXINSTVALFMV
FYHINITKTKPLSLMNKSPTLMLTTILVLLSLGGLPPLTGFAPKWLVVOELIMHNNI
MMAATTLAIMALLNLFYMRIIYSSLTITTFPTNNKXHWYKQPMKTPLSPLPTILST
ALLPLTPMFTLG"
gene 5013. .5081
/gene="tRNA-Trp"
tRNA 5013. .5081
/gene="tRNA-Trp"
/product="tRNA-Trp"
/note="codon recognized: UGA"
/anticodon=(pos:5043. .5045,aa:Trp)
gene complement(5085. .5151)
/gene="tRNA-Ala"
complement(5085. .5151)
/gene="tRNA-Ala"
/product="tRNA-Ala"
/note="codon recognized: GCA"
/anticodon=(pos:5119. .5121,aa:Ala)
tRNA complement(5152. .5524)
/gene="tRNA-Asn"
complement(5152. .5524)
/gene="tRNA-Asn"
/product="tRNA-Asn"
/note="codon recognized: AAC"
/anticodon=(pos:5189. .5191,aa:Asn)
gene complement(5257. .5317)
/gene="tRNA-Cys"
complement(5257. .5317)
/gene="tRNA-Cys"
/product="tRNA-Cys"
/note="codon recognized: UUU"
/anticodon=(pos:5287. .5289,aa:Cys)
tRNA complement(5318. .5383)
/gene="tRNA-Tyr"
complement(5318. .5383)
/gene="tRNA-Tyr"
/product="tRNA-Tyr"
/note="codon recognized: UUU"
/anticodon=(pos:5350. .5352,aa:Tyr)
gene 5384. .6925
/gene="col1"
CDS 5384. .6925
/gene="col1"
/codon_start=1
/transl_table=2
/product="cytochrome c oxidase subunit 1"
/protein_id="CAC88012.1"
/db_xref="GI:17932767"
/translation="MFINRWLFSTNHKDIGTLLYLLFGAWAGMAGTALSILIRSELGQP
GSLGDDQIYNVITYAHAFVMIEFMVMPIMIGFGNMLVPLMIGADPMAFPRMNNMSF
WLLPSPFLLLLVSVTEAGAGTGTVYPLLAGNLAHAGASVDLAIFSLLHAGVSSILG
AINFITTIINMKPRAMSQYQTPLFVMSVLYTAVULLLSLPVLAAGITMLTDRNLNTT
FPDPAGGDPILYOHLEFWFEGHPEVYILILPFGIISHIVTYISGKKEPFGYMGWMA
MMSIGFLGFIWVAHNMFTVGMVDVTRAYFTSATMIIAIPTVKVFSSMLATLHGDIKW
TPPMLWALGFIFLFTVGGLTGIVLANSLLDILLHDITYVVAHFHYLSMGAVFAIMGG
FVHWFPPLLSGFTLHTWAKVHFTLMFVGVNLTFFPQHFLGAGMPRAYSDDPAYTLM
NAISSIGSFISLTAVVMVIEFMVEAFASKREVLTVELTSTNIEWLHGCCPPYHTFEEP
VIYKI"
gene complement(6937. .6997)
/gene="tRNA-Ser (UCN)"

```
trna      complement(6937. .6997)
          /gene="tRNA-Ser (UCN)"
          /product="tRNA-Ser"
          /note="codon recognized: UCA"
          /anticodon=(pos:6963. .6965,aa:Ser)
          6997. .7068
          /gene="tRNA-Asp"
          6997. .7068
          /gene="tRNA-Asp"
          /product="tRNA-Asp"

Query Match      93.3%; Score 14; DB 4; Length 16360;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 CCTCTTGGATCGGC 15
        |||
        9647 CCTCTTGGATCGGC 9660

RESULT 11
AP000389      55739 bp      DNA      linear      PLN 27-DEC-2000
LOCUS      Arabidopsis thaliana genomic DNA, chromosome 3, P1 clone:MSJ3.
DEFINITION      AP000389 BA000014
ACCESSION      AP000389.1 GI:5672589
VERSION
KEYWORDS
SOURCE
ORGANISM      Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui P1
clone:MSJ3.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (sites)
kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.
Structural analysis of Arabidopsis thaliana chromosome 3. II.
Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
and BAC clones
DNA Res. 7 (3), 217-221 (2000)
20363099
2 (bases 1 to 55739)
kaneko,T., Katoh,T., Sato,S., Nakamura,Y., Asamizu,E. and Tabata,S.
Direct Submission
Submitted (28-JUL-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research; 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail:ynakamu@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MSJ3
Genes with similarity to proteins in the databases are described in
'product' or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
http://compbio.ornl.gov/Grail-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremli1.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is MVA11 and the 3' clone is MQP15.

FEATURES
source
1. .55739
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"

CDS
/clone="MSJ3"
/clone_lib="Mitsui P1"
join(462. .735,1345. .1373)
/note="gb|AAC61810.1
gene_id:MSJ3.1
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01420.1"
/db_xref="GI:11994245"
/translation="MDPNKAEISRLVLPQDLIGEIVAKTGAKYAEDYHNCILSCKEL
GASANDERYLKTNLAPLVKKPLSCRKHLIMKKCIANNNDPAHYIKANSRNVK"
complement(join(2420. .4168,4328. .5023,5138. .5459))
/note="contains similarity to Arabidopsis thaliana
retrotransposon Athila
gene_id:MSJ3.2"
/pseudo
/codon_start=1
/evidence=not_experimental
6784. .8561
/note="gene_id:MSJ3.3
unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
8666. .9013
/note="gb|AAF24530.1
gene_id:MSJ3.4
similar to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
join(13427. .13562,13605. .13769,14564. .15219,15302. .15409,
15500. .15556,15681. .15725,15808. .16049,16127. .16235,
16688. .16859,16997. .17124,17227. .17382)
/note="gb|AAC97213.1
gene_id:MSJ3.5
strong similarity to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
join(19136. .20065,20144. .20368,20467. .20550,20807. .20926,
21101. .21220,21809. .21955,22036. .22182,22263. .22353,
22432. .22550,22624. .22794,22894. .22941)
/note="gb|AAD25853.1
gene_id:MSJ3.6
strong similarity to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01421.1"
/db_xref="GI:11994246"
/translation="MSTKKLFFPKRLFQPGSEVDNITRVSKHSSFKIVTVRKALPEI
EFNTLMDTFLGPIVEFAEMDLAFSSHIVHLLQRRILTNDELWFVADQPMRFSLE
FITTTGLPMDDASGSSPNVKKRKKKWKIKDNQTLKGGKKDWINKEMTFLDLTYVL
EKRSKMQPEERLRLVGSLLVEGILLARNPTTMLPIENLRATDFEKFCKPWAQIVY
SYLVTEVRRIGNDVYLLRKQYGLFGFVQSIQIWLSSVQKLGKFGTRDESSGNIPLI
LOWRSTSSPSAKAITMIADGDKVHFGFVYLQFKLFTDFKRLIKDLIDVFDLVI
KGYRMSKKDWQGWIQVEQGHQGRIDHVLKPAIETVHQSRRKSSMSFGKRRTRRG
RKKEQEEERQEEEGKEELEKVEYRGDETEKQEI PKQDEMEGEKEEKEEGKEE
EKVEYRGDETEKQEI PKQGNEMEVEEKEKEEKEEKEEKEEKEEKEEKEEKEE
TEDEAVVILSDDENGTAPTEKESQPKETEETEVPKKEENVEHDEHDETEDEAVI
LSDEDENGTAPTEKESQPKETEETEVRETCKDDEVDNITFLFQEEITOGSSLOT
PLTPVMLSQEWEEIDLKRVKWKAKNLIRDLSSLEELIWDSSKQKVEIWDWDEK
KYLALKKATLLIHPDKLLRAHPEVAYLVEQICKIIMVEKDATKAKMDLVL"
complement(join(23766. .24009,24128. .25699,25824. .25874,
26025. .26075,26157. .26453))
/note="gene_id:MSJ3.7"
/pseudo
/codon_start=1
/evidence=not_experimental
/product="Mutator-like transposase"
33238. .33464
/note="contains similarity to En/Spm-like transposon
```


protein
gene_id:MSJ3.8"
/pseudo
/codon_start=1
/evidence=not_experimental
39943. .40386
/note="gene_id:MSJ3.9
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01422.1"
/db_xref="GI:11994247"
/translation="MAEEFHLPLNLTPEPDQPTTEDEFEADTTLIDEXLYRLHYLDSR
SNHGVTFGSYDEIFGLVLGNSSSCTRWLNAGEELPVVEFTAEEMERGLVCAICRE
ELANERLSELPCRHYHKECISNWLSNRNTCPLCRHNVELPNHG"
42049. .47107
/note="gb|AAD15532.1
gene_id:MSJ3.10
strong similarity to unknown protein"
/pseudo
/codon_start=1
/evidence=not_experimental
complement(join(47610. .47819,47908. .48078,48155. .48202))
/note="gb|AAD23021.1
gene_id:MSJ3.11
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01423.1"
/db_xref="GI:11994248"
/translation="MAQRSLFQGESSDDEELDMETRMTRYGENRRRDKGVPIED
CNAKVVATSLDPVTTEKLEFSCPYEISDGPGGCFKRWTVALCDEFDMKEETSE
MKKDLAANKRVESQAEKIFLMEKKFETLEKTESLNKYL"
join(49962. .50126,50196. .50501)
/note="gb|AAC69115.1
gene_id:MSJ3.12
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01424.1"
/db_xref="GI:11994249"
/translation="MSSNSSENSIASTOSVSRWGRGLPSKCHGCLDVITYTSYAKNPG
RPFRCPTRKDDHLFKWVEDGTYEVVDSPKLSYMDSELGSAKSEVAVELEALQAMV
KEJKEKVMRSERENQIKWLKLCFCFLAFTIIVFGILMPKINERQLFLGY"
complement(join(51037. .51156,51256. .51486))
/note="contains similarity to En/Spm-like transposon
protein
gene_id:MSJ3.13"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01425.1"
/db_xref="GI:11994250"
/translation="WPHITGOLFRLDVEVVLTAAPRADLPVLNLGLGTWVDKTYTW
KFHQDLAVAIGDILQRFNHAMPYQIMIPVFKRMHTWDSRHLLTWTFQFNLTVAFTLY
RDNMRALKRDRGV"
join(52302. .52436,52529. .52793,52867. .52969,53069. .53312)
/note="gene_id:MSJ3.14"
/codon_start=1
/evidence=not_experimental
/product="replication protein A1-like"
/protein_id="BAB01426.1"
/db_xref="GI:11994251"
/translation="MGYDGLRDLNPCICTCWTICVMVFRMYISFPTDNSFELWVVEDE
WGQIEAIIIESRFQRYMDLEENQWKEIRTFGVIENGCFVRSTHKYMIELMATEV
THSDPRETVLENFTPFVDYIIEPTAHTEMLVDLGVGLVGCCLRNITEDGWMEDGFL
TFEIMDKYERLKECEAFGELAVDLHRRFSFRNDSIVIDNIFVALLTWKIFRYHTGLK
HVKFONYAGISDVTVDPNIEVELEKMLE"
BASE COUNT 18008 a 10125 c 10679 g 16927 t
ORIGIN
Query Match 93.3%; Score 14; DB 8; Length 55739;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGC 15
|||||
Db 30106 CCTCTTGATCGC 30093
RESULT 12
AC123621/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
1 (bases 1 to 67079)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-74D13
Unpublished
2 (bases 1 to 67079)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Marquis,N.,
Mathews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Topham,K., Travers,M., Travis,N., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23301
Center clone name: 74_D_13

* NOTE: This record contains 85 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 698: contig of 698 bp in length
* 699 798: gap of 100 bp
* 799 1492: contig of 694 bp in length

[illegible]

*	30064	30762:	contig of 699 bp	in length
*	30763	30862: gap of	100 bp	
*	30863	31540: contig of 678 bp	in length	
*	31541	31640: gap of	100 bp	
*	31641	32339: contig of 699 bp	in length	
*	32340	32439: gap of	100 bp	
*	32440	33128: contig of 689 bp	in length	
*	33129	33228: gap of	100 bp	
*	33229	33873: contig of 645 bp	in length	
*	33874	33973: gap of	100 bp	
*	33974	34656: contig of 683 bp	in length	
*	34657	34756: gap of	100 bp	
*	34757	35450: contig of 694 bp	in length	
*	35451	35550: gap of	100 bp	
*	35551	36231: contig of 681 bp	in length	
*	36232	36331: gap of	100 bp	
*	36332	37027: contig of 696 bp	in length	
*	37028	37127: gap of	100 bp	
*	37128	37828: contig of 701 bp	in length	
*	37828	37928: gap of	100 bp	
*	37929	38621: contig of 693 bp	in length	
*	38622	38721: gap of	100 bp	
*	38722	39407: contig of 686 bp	in length	
*	39408	39507: gap of	100 bp	
*	39508	40214: contig of 707 bp	in length	
*	40215	40314: gap of	100 bp	
*	40315	41005: contig of 691 bp	in length	
*	41006	41105: gap of	100 bp	
*	41106	41802: contig of 697 bp	in length	
*	41803	41902: gap of	100 bp	
*	41903	42601: contig of 699 bp	in length	
*	42602	42701: gap of	100 bp	
*	42702	43393: contig of 692 bp	in length	
*	43394	43493: gap of	100 bp	
*	43494	44185: contig of 692 bp	in length	
*	44186	44285: gap of	100 bp	
*	44286	44955: contig of 670 bp	in length	
*	44956	45055: gap of	100 bp	
*	45056	45742: contig of 687 bp	in length	
*	45743	45842: gap of	100 bp	
*	45843	46535: contig of 693 bp	in length	
*	46536	46635: gap of	100 bp	
*	46636	47332: contig of 697 bp	in length	
*	47333	47432: gap of	100 bp	
*	47433	48126: contig of 694 bp	in length	
*	48127	48226: gap of	100 bp	
*	48227	48903: contig of 677 bp	in length	
*	48904	49003: gap of	100 bp	
*	49004	49705: contig of 702 bp	in length	
*	49706	49805: gap of	100 bp	
*	49806	50476: contig of 671 bp	in length	
*	50477	50576: gap of	100 bp	
*	50577	51270: contig of 694 bp	in length	
*	51271	51370: gap of	100 bp	
*	51371	52053: contig of 683 bp	in length	
*	52054	52153: gap of	100 bp	
*	52154	52843: contig of 690 bp	in length	
*	52844	52943: gap of	100 bp	
*	52944	53618: contig of 675 bp	in length	
*	53619	53718: gap of	100 bp	
*	53719	54402: contig of 684 bp	in length	
*	54403	54502: gap of	100 bp	

Query Match	93.3%;	Score 14;	DB 2;	Length 67079;
Best Local Similarity	100.0%;	Pred. No. 1.5e+03;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0.

QY	2	CCTCTTGATCGGC	15
Db	64842	CCTCTTGATCGGC	64829

RESULT 13

AL592300/c	AL592300	68054 bp	DNA	linear	PRI 28-FEB-2002
LOCUS	Human DNA sequence from clone RP11-294K24 on chromosome 1, complete sequence.				
DEFINITION					
ACCESSION	AL592300				
VERSION	AL592300.16	GI:19068251			
KEYWORDS	HTG.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 68054)				
TITLE	Pearce, A.				
JOURNAL	Direct Submission				
COMMENT	Submitted (28-FEB-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Mar 1, 2002 this sequence version replaced gi:16973133. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TrEMBL; Wp:, WormPEP; information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr1 RP11-294K24 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see http://www.chori.org/bacpac/home.htm VECTOR: PBACE3.6 IMPORTANT: This sequence is not the entire insert of clone RP11-294K24. It may be shorter because we sequence overlapping sections only once, except for a short overlap. The true left end of clone RP11-294K24 is at 1 in this sequence. The true left end of clone RP11-572A16 is at 66055 in this sequence.				
FEATURES	Location/Qualifiers				
Source	1.68054				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/chromosome="1"				
	/clone="RP11-294K24"				
	/clone_1lb="RPCI-11.2"				
BASE COUNT	14580 a 18489 c 18158 g 16827 t				
ORIGIN					
Query Match	93.3%;	Score 14;	DB 9;	Length 68054;	
Best Local Similarity	100.0%;	Pred. No. 1.5e+03;			
Matches	14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	2 CCTCTTGATCGGC 15				
Db	6184 CCTCTTGATCGGC 6171				
RESULT 14					
LOCUS	AC100880	69928 bp	DNA	linear	HTG 23-NOV-2001
DEFINITION	Mus musculus clone RP23-6716, LOW-PASS SEQUENCE SAMPLING.				

ACCESSION	ACI00880
VERSION	ACI00880.1 GI:17059654
KEYWORDS	HTG; HTGS_PHASE0.
SOURCE	Mus musculus.
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	Birren,B., Linton,L., Nusbaum,C. and Lander,E. 1 (bases 1 to 69928)
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE	Mus musculus, clone RP23-6716
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 69928) Birren,B., Linton,L., Nusbaum,C., Lander,E., All,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgatter,B., Brown,A., Camarata,J., Campiolano,A., Chang,J., Chazarov,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Hearford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R., Landers,T., Lehoczyk,J., Levine,R., Liu,G., Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J., Menues,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and zody,M.
TITLE	Direct Submission
JOURNAL	Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html
	----- Genome Center ----- Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information ----- Center project name: LI4490 Center clone name: 67_1_6 -----
	* NOTE: This record contains 82 individual * sequencing reads that have not been assembled into * contigs. Runs of N are used to separate the reads * and the order in which they appear is completely * arbitrary. Low-pass sequence sampling is useful for * identifying clones that may be gene-rich and allows * overlap relationships among clones to be deduced. * However, it should not be assumed that this clone * will be sequenced to completion. In the event that * the record is updated, the accession number will * be preserved. * 1 763: contig of 763 bp in length * 764 863: gap of 100 bp * 864 1613: contig of 750 bp in length * 1614 1713: gap of 100 bp * 1714 2478: contig of 765 bp in length * 2479 2578: gap of 100 bp * 2579 3313: contig of 735 bp in length * 3314 3413: gap of 100 bp * 3414 4122: contig of 709 bp in length * 4123 4222: gap of 100 bp * 4223 4981: contig of 759 bp in length * 4982 5081: gap of 100 bp * 5082 5824: contig of 743 bp in length

```
* 5825 5924: gap of 100 bp
* 5925 6695: contig of 771 bp in length
* 6696 6795: gap of 100 bp
* 6796 7561: contig of 766 bp in length
* 7562 7661: gap of 100 bp
* 7662 8413: contig of 752 bp in length
* 8414 8513: gap of 100 bp
* 8514 9262: contig of 749 bp in length
* 9263 9362: gap of 100 bp
* 9363 10114: contig of 752 bp in length
* 10115 10214: gap of 100 bp
* 10215 10958: contig of 744 bp in length
* 10959 11058: gap of 100 bp
* 11059 11816: contig of 758 bp in length
* 11817 11916: gap of 100 bp
* 11917 12656: contig of 740 bp in length
* 12657 12756: gap of 100 bp
* 12757 13505: contig of 749 bp in length
* 13506 13605: gap of 100 bp
* 13606 14372: contig of 767 bp in length
* 14373 14472: gap of 100 bp
* 14473 15241: contig of 769 bp in length
* 15242 15341: gap of 100 bp
* 15342 16114: contig of 773 bp in length
* 16115 16214: gap of 100 bp
* 16215 16964: contig of 750 bp in length
* 16965 17064: gap of 100 bp
* 17065 17805: contig of 741 bp in length
* 17806 17905: gap of 100 bp
* 17906 18670: contig of 765 bp in length
* 18671 18770: gap of 100 bp
* 18771 19523: contig of 753 bp in length
* 19524 19623: gap of 100 bp
* 19624 20375: contig of 752 bp in length
* 20376 20475: gap of 100 bp
* 20476 21236: contig of 761 bp in length
* 21237 21336: gap of 100 bp
* 21337 22069: contig of 733 bp in length
* 22070 22169: gap of 100 bp
* 22170 22897: contig of 728 bp in length
* 22898 22997: gap of 100 bp
* 22998 23767: contig of 770 bp in length
* 23768 23867: gap of 100 bp
* 23868 24628: contig of 761 bp in length
* 24629 24728: gap of 100 bp
* 24729 25482: contig of 754 bp in length
* 25483 25582: gap of 100 bp
* 25583 26357: contig of 775 bp in length
* 26358 26457: gap of 100 bp
* 26458 27202: contig of 745 bp in length
* 27203 27302: gap of 100 bp
* 27303 28046: contig of 744 bp in length
* 28047 28146: gap of 100 bp
* 28147 28878: contig of 732 bp in length
* 28879 28978: gap of 100 bp
* 28979 29733: contig of 755 bp in length
* 29734 29833: gap of 100 bp
* 29834 30596: contig of 763 bp in length
* 30597 30696: gap of 100 bp
* 30697 31462: contig of 766 bp in length
* 31463 31562: gap of 100 bp
* 31563 32332: contig of 770 bp in length
* 32333 32432: gap of 100 bp
* 32433 33169: contig of 737 bp in length
* 33170 33269: gap of 100 bp
* 33270 34011: contig of 742 bp in length
* 34012 34111: gap of 100 bp
* 34112 34864: contig of 753 bp in length
* 34865 34964: gap of 100 bp
* 34965 35718: contig of 754 bp in length
* 35719 35818: gap of 100 bp
* 35819 36574: contig of 756 bp in length
* 36575 36674: gap of 100 bp
```

```
* 36675 37434: contig of 760 bp in length
* 37435 37534: gap of 100 bp
* 37535 38304: contig of 770 bp in length
* 38305 38404: gap of 100 bp
* 38405 39149: contig of 745 bp in length
* 39150 39249: gap of 100 bp
* 39250 39992: contig of 743 bp in length
* 39993 40092: gap of 100 bp
* 40093 40857: contig of 765 bp in length
* 40858 40957: gap of 100 bp
* 40958 41722: contig of 765 bp in length
* 41723 41822: gap of 100 bp
* 41823 42585: contig of 763 bp in length
* 42586 42685: gap of 100 bp
* 42686 43452: contig of 767 bp in length
* 43453 43552: gap of 100 bp
* 43553 44311: contig of 759 bp in length
* 44312 44411: gap of 100 bp
* 44412 45161: contig of 750 bp in length
* 45162 45261: gap of 100 bp
* 45262 46021: contig of 760 bp in length
* 46022 46121: gap of 100 bp
* 46122 46870: contig of 749 bp in length
* 46871 46970: gap of 100 bp
* 46971 47702: contig of 732 bp in length
* 47703 47802: gap of 100 bp
* 47803 48542: contig of 740 bp in length
* 48543 48642: gap of 100 bp
* 48643 49399: contig of 757 bp in length
* 49400 49499: gap of 100 bp
* 49500 50261: contig of 762 bp in length
* 50262 50361: gap of 100 bp
* 50362 51130: contig of 769 bp in length
* 51131 51230: gap of 100 bp
* 51231 51988: contig of 758 bp in length
* 51989 52088: gap of 100 bp
* 52089 52859: contig of 771 bp in length
* 52860 52959: gap of 100 bp
* 52960 53696: contig of 737 bp in length
* 53697 53796: gap of 100 bp
* 53797 54546: contig of 750 bp in length
* 54547 54646: gap of 100 bp
* 54647 55408: contig of 762 bp in length
* 55409 55508: gap of 100 bp
* 55509 56259: contig of 751 bp in length
* 56260 56359: gap of 100 bp
* 56360 57092: contig of 733 bp in length
* 57093 57192: gap of 100 bp
* 57193 57911: contig of 719 bp in length
* 57912 58011: gap of 100 bp
* 58012 58778: contig of 767 bp in length
* 58779 58878: gap of 100 bp
```

Query Match 93.3%; Score 14; DB 2; Length 69928;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGG 14
DB 27437 ACCTCTTGATCGG 27424

RESULT 15

AL137862/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AL137862 73509 bp DNA linear HTG 20-OCT-2001
Homo sapiens chromosome 11 clone RP5-1002E13 map p12-14.1, ***
SEQUENCING IN PROGRESS ***, in ordered pieces.
AL137862 GI:16304424
HTG; HTGS_PHASE2; HTGS_CANCELLED.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS 1 (bases 1 to 73509)
TITLE Wallis, J.
JOURNAL Direct Submission

COMMENT Submitted (16-Oct-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 21, 2001 this sequence version replaced gi:16214503.
----- Genome Center

Center: Sanger Centre
Center code: SC

Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

----- Project Information
Center project name: dj1002E13

----- Summary Statistics
Assembly program: XGAP4; version 4.5

Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Consensus quality: 72684 bases at least Q40
Consensus quality: 73024 bases at least Q30

Consensus quality: 73236 bases at least Q20
Insert size: 73509; sum-of-contigs

Insert size: 93120; 6.4% error; agarose-fp
Quality coverage: 4.88x in Q20 bases; sum-of-contigs Quality
coverage: 4.40x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.

* This sequence will be replaced
* by the finished sequence as soon as it is available and

* the accession number will be preserved.

FEATURES

source

Location/Qualifiers

1..73509

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="11"

/map="p12-14.1"

/clone="RP5-1002E13"

/clone_lib="RPC1-5"

1..73509

/note="assembly_fragment:00544"

BASE COUNT 23859 a 13784 c 13831 g 22034 t 1 others

ORIGIN

Query Match 93.3%; Score 14; DB 2; Length 73509;

Best Local Similarity 100.0%; Pred. No. 1.5e+03;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGG 14
|||||

DB 56207 ACCTCTTGATCGG 56194

Search completed: June 26, 2003, 03:47:27
Job time : 429.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds

(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-1

Perfect score: 15

Sequence: 1 acccttgatcggc 15

Scoring table: IDENTITY_NUC

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*

24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	20	AAAX32387	Abl variable heavy
2	14	93.3	1443	19	AAV27174	G-protein coupled
3	14	93.3	1446	21	AAA30638	Human G protein-co
4	14	93.3	1446	21	AAA30731	DNA encoding human
5	14	93.3	1616	22	AAH98401	Human EST-derived
6	14	93.3	1626	19	AAV27175	G-protein coupled
7	14	93.3	1728	20	AAZ11093	Human mucilage cod
8	14	93.3	1850	21	AAZ36408	cDNA encoding a G-
9	14	93.3	2411	20	AAZ00860	Human brain G-prot

C	10	14	93.3	1038602	20	AAZ01425
C	11	13.4	89.3	329	22	AAAS34099
C	12	13.4	89.3	360	15	AAQ66413
C	13	13.4	89.3	366	22	AAAS03486
C	14	13.4	89.3	375	22	AAH66453
C	15	13.4	89.3	387	22	AAAS03465
C	16	13.4	89.3	390	24	ABL78599
C	17	13.4	89.3	466	21	AAA91626
C	18	13.4	89.3	538	24	ABK64653
C	19	13.4	89.3	615	23	ABL22291
C	20	13.4	89.3	680	22	AAH07127
C	21	13.4	89.3	819	21	AAF21556
C	22	13.4	89.3	819	21	AAF21557
C	23	13.4	89.3	819	21	AAF21559
C	24	13.4	89.3	819	21	AAA81271
C	25	13.4	89.3	819	21	AAA81272
C	26	13.4	89.3	819	21	AAA81274
C	27	13.4	89.3	819	21	AAZ53896
C	28	13.4	89.3	819	21	AAZ53898
C	29	13.4	89.3	819	21	AAZ53899
C	30	13.4	89.3	819	21	AAZ54588
C	31	13.4	89.3	819	21	AAZ54589
C	32	13.4	89.3	819	21	AAZ54591
C	33	13.4	89.3	909	23	AAZ54720
C	34	13.4	89.3	909	23	AAZ54530
C	35	13.4	89.3	1177	20	AAV64416
C	36	13.4	89.3	1299	21	AAF07517
C	37	13.4	89.3	1350	22	AAH65217
C	38	13.4	89.3	1836	22	AAH17969
C	39	13.4	89.3	2549	23	ABL22284
C	40	13.4	89.3	2844	23	ABL22290
C	41	13.4	89.3	2944	23	ABL09850
C	42	13.4	89.3	3004	22	AAH15779
C	43	13.4	89.3	3775	23	ABL12027
C	44	13.4	89.3	6199	23	ABL12026
C	45	13.4	89.3	6670	23	ABL25222

ALIGNMENTS

RESULT 1	AAAX32387	standard; DNA: 15 BP.	Complete genome se
ID	AAAX32387		Human cDNA encodin
AC	AAAX32387;		alca promoter. As
XX			DNA encoding anti-
XX			C glutamicum codin
DT	17-JUN-1999	(first entry)	DNA encoding anti-
XX			Human ovarian canc
DE	Abl variable heavy (VH) chain CDR1 encoding DNA.		Caenorhabditis ele
XX			Human benign prost
KW	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;		Drosophila melanog
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;		Human cDNA clone (
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;		N. meningitidis pa
KW	myelodysplasia; myelotoxic chemotherapy; leukaemia; tumour; MMSK; CDR;		N. meningitidis pa
KW	neuromuscular; muscular dystrophy; complementarity determining region;		N. meningitidis pa
XX	variable heavy chain; variable light chain; VH; VL; SS.		N. meningitidis pa
OS	Homo sapiens.		N. meningitidis pa
XX			N. meningitidis pa
PN	WO9910494-A2.		N. meningitidis pa
XX			N. meningitidis pa
PD	04-MAR-1999.		N. meningitidis pa
XX			N. meningitidis pa
PF	21-AUG-1998;	98WO-US17364.	N. meningitidis pa
XX			N. meningitidis pa
PR	25-AUG-1997;	97US-0918148.	N. meningitidis pa
XX			N. meningitidis pa
PA	(GETH) GENENTECH INC.		N. meningitidis pa
XX			N. meningitidis pa
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;		N. meningitidis pa
XX			N. meningitidis pa
DR	WPI; 1999-204666/17.		N. meningitidis pa

CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
 CC from the organism of interest. They can be used in diagnostics,
 CC forensics, gene mapping, identification of mutations, to assess
 CC biodiversity and for nutritional purposes. The present sequence is a CDNA
 CC of the invention.

XX
 SQ Sequence 1616 BP; 269 A; 526 C; 480 G; 341 T; 0 other;

Query Match 93.3%; Score 14; DB 22; Length 1616;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 Db 166 CCTCTTGATCGGC 153

RESULT 6
 AAV27175/c
 ID AAV27175 standard; cDNA; 1626 BP.

XX AC AAV27175;

XX DT 17-SEP-1998 (first entry)

XX DE G-protein coupled receptor, long form, coding sequence.

XX KW G-protein coupled receptor; gene therapy; abnormality detection;
 long form; human; ds.

XX OS Homo sapiens.

XX PN EP845529-A2.

XX PD 03-JUN-1998.

XX PF 27-OCT-1997; 97EP-0308562.

XX PR 29-OCT-1996; 96JP-0286823.

XX PA (TAKE) TAKEDA CHEM IND LTD.

XX PI Fukusumi S, Hinuma S, Kawamata Y;

XX DR WPI: 1998-288746/26.

XX DR P-PSDB; AAW55030.

XX PT New human G-protein coupled receptor protein - and corresponding
 DNA, ligands, antibodies, etc

XX PS Claim 6; Page 50-51; 65pp; English.

CC This sequence encodes a human G-protein coupled receptor of the
 CC invention. The protein or cells expressing the DNA encoding it can be
 CC used to screen for agonists or antagonists of the receptor, which can be
 CC used as drugs for treating various diseases (none disclosed). The DNA can
 CC also be used for practice drug design based on comparisons with
 CC structurally analogous ligands and receptors. DNA encoding the protein
 CC can be used for gene therapy for diseases caused by a deficiency of the
 CC receptor. The DNA can also be used to detect abnormalities in the gene
 CC encoding the receptor. The protein or fragment can be used to determine
 CC levels of receptor ligands in vivo. The antibody can be used in assays to
 CC detect the protein.

XX SQ Sequence 1626 BP; 272 A; 540 C; 484 G; 330 T; 0 other;

Query Match 93.3%; Score 14; DB 19; Length 1626;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 Db 321 CCTCTTGATCGGC 308

RESULT 7
 AA211093/c
 ID AA211093 standard; cDNA; 1728 BP.

XX AC AA211093;

XX DT 03-NOV-1999 (first entry)

XX DE Human mucilage coding sequence.

XX KW Mucilage; human; G-protein-coupled receptor; infection; cancer; diabetes;
 obesity; anorexia; Parkinson's disease; hypotension; hypertension;
 osteoporosis; stroke; myocardial infarction; asthma; allergy; therapy;
 neurological disorder; dyskinesia; diagnosis; ss.

XX OS Homo sapiens.

XX FH Key location/Qualifiers
 FT CDS 150..1595
 FT /*tag= a
 FT /product= mucilage

XX PN WO9942485-A1.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03320.

XX PR 08-MAY-1998; 98US-0075468.

XX PR 20-FEB-1998; 98US-0075307.

XX PA (SMIK) SMITHKLINE BEECHAM CORP.

XX PI Elshourbagy N, Vawter L;

XX DR WPI: 1999-527457/44.

XX DR P-PSDB; AAY32886.

XX PT New mucilage receptor polypeptide for treatment, prevention and
 diagnosis of e.g. infections and cancer

XX PS Claim 2; Page 34; 50pp; English.

CC This sequence encodes the mucilage protein of the invention. Mucilage is
 CC a G-protein-coupled receptor. Mucilage, or the nucleic acid encoding it,
 CC are used to treat conditions requiring increased activity/expression of
 CC mucilage, while antagonists (or nucleic acid encoding them or competitive
 CC peptides) of the protein are used where the opposite effect is required.
 CC Typical of very many diseases that can be treated are infections (viral,
 CC bacterial, fungal, protozoal, particularly human immune deficiency
 CC virus-1 or HIV-2); pain; cancer; diabetes; obesity; anorexia; Parkinson's
 CC disease; hypotension or hypertension; osteoporosis; stroke; myocardial
 CC infarction; asthma; allergy; neurological disorders; dyskinesia etc.
 CC Detecting mutations in the DNA and/or determining the presence or amount
 CC of the protein is used to diagnose these diseases, or susceptibility to
 CC them. The mucilage protein is also used to screen for specific
 CC (ant)agonists, which are potential therapeutic agents; to raise
 CC antibodies (Ab); in protective vaccines; to isolate cognate receptors;
 CC and in structure-based drug design. The mucilage coding sequences are
 CC used to produce recombinant mucilage; as a source of primers and probes
 CC for isolating related genes or detecting mutations; for chromosome
 CC identification; and as source of antagonistic oligonucleotides. The Ab
 CC are used to identify or isolate mucilage-expressing cells, for affinity
 CC purification of mucilage and therapeutically.

XX SQ Sequence 1728 BP; 284 A; 569 C; 513 G; 362 T; 0 other;

Query Match 93.3%; Score 14; DB 20; Length 1728;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
 |||||
 Db 287 CCTCTTGATCGGC 274

RESULT 8

AAZ36408/c
 ID AAZ36408 standard; cDNA; 1850 BP.

XX AAZ36408;

XX 22-FEB-2000 (first entry)

XX cDNA encoding a G-protein coupled receptor designated HG01.

XX Human; G-protein coupled receptor; HG01; endothelin receptor;

KW pathological state; gene therapy; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 84..1529

FT /*tag= a

FT /product= "G-protein coupled receptor HG01"

FT /transl_except= (pos: 186..188, aa: Ala)

FT sig_peptide 84..158

FT /*tag= b

FT mat_peptide 159..1526

FT /*tag= c

FT /note= "specifically claimed in claim 2"

XX WO9960153-A2.

XX 25-NOV-1999.

XX 17-MAY-1999; 99WO-US10808.

XX 21-MAY-1998; 98US-0086294.

XX (MERI) MERCK & CO INC.

XX Liu Q;

XX WPI: 2000-053303/04.

DR P-PSDB; AAY53779.

XX Novel receptor protein, useful for identifying antagonists and agonists for drug development

XX Claim 2; Fig 1A; 33pp; English.

XX The present sequence encodes a human G-protein coupled receptor, designated HG01. HG01 is highly expressed in the brain, and is thought to be related to the endothelin receptors. The HG01 cDNA sequence was isolated from a human foetal brain library. HG01 can be used to identify binding substances, antagonists and agonists, which may be useful pharmacologically to modulate HG01 activity. G-protein coupled receptors (GPCRs) are involved in many pathological states, and modulators of GPCR activity are therefore useful to treat such conditions. HG01 (or fragments) are also useful to produce antibodies. The polynucleotides can be used in gene therapy, especially to treat conditions requiring increased HG01.

XX Sequence 1850 BP; 305 A; 614 C; 540 G; 390 T; 1 other;

XX Query Match

XX Best Local Similarity 93.3%; Score 14; DB 21; length 1850;

XX Matches 14; Conservativity 100.0%; Pred. NO. 2.3e+02; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
 |||||

Db 221 CCTCTTGATCGGC 208

RESULT 9

AAZ00860/c
 ID AAZ00860 standard; cDNA to mRNA; 2411 BP.

XX AAZ00860;

XX 11-OCT-1999 (first entry)

XX Human brain G-protein coupled receptor cDNA.

XX G-protein coupled receptor; human; brain; anti-neurodegeneration; pain;

KW antitumour; analgesic; anti-epileptic; anti-addictive; anti-obesity;

KW anti-anorexic; anti-bulimic; cerebroprotective; gene therapy; diagnostic;

KW detection; disease; central nervous system; peripheral nervous system;

KW Alzheimer's disease; Parkinson's disease; brain tumour; schizophrenia;

KW epilepsy; drug addiction; obesity; anorexia; bulimia; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

FT 5'UTR 1..19

FT /*tag= a

FT CDS 20..1465

FT /*tag= b

FT /product= "G-protein coupled receptor"

FT 3'UTR 1466..2411

FT /*tag= c

XX DE19805351-A1.

XX 12-AUG-1999.

XX 11-FEB-1998; 98DE-1005351.

XX 11-FEB-1998; 98DE-1005351.

XX (BADI) BASF AG.

XX Kroeger B, Otterbach B;

XX WPI: 1999-445442/38.

DR P-PSDB; AAY25969.

XX New human G-protein coupled receptor from brain tissue, used to treat nervous system disorders, e.g. Alzheimer's disease, eating disorders and as cerebral protectant

XX Claim 5; Page 6-8; 12pp; German.

XX This invention describes a novel G-protein coupled receptor isolated from human brain tissue which has anti-neurodegeneration, antitumour, analgesic, anti-epileptic, anti-addictive, anti-obesity, anti-anorexic, anti-bulimic and cerebroprotective activity. Nucleic acid (II) that encodes (I), or its complement, is used (I) in gene therapy of (I)-deficiency states, (2) as probes and primers for diagnostic detection of (II) in usual hybridization/amplification assays, (3) for recombinant expression of (I), (4) to isolate the corresponding genomic sequence, (5) to design ribozymes and antisense sequences for treating overexpression of (I), (6) to produce transgenic animals for studying the (patho)physiology of (I) and (7) to detect mutations in (I)-encoding genes. (1) is used (1) to raise specific antibodies (Ab), (2) to screen for specific ligands, agonists and antagonists, (potentially useful as therapeutic agents) and (3) to treat (I)-deficiency conditions. Abs to (I) are used to detect/determine (I) in standard immunoassays, particularly for diagnosis. (I) is associated with chronic and acute diseases of the central and peripheral nervous systems (e.g. Alzheimer's or Parkinson's diseases, brain tumours, pain, schizophrenia, epilepsy, drug addiction etc.), diseases that include a nervous component (e.g. obesity, anorexia and bulimia). This sequence encodes the human brain G-protein coupled receptor described in the invention.

XX Sequence 2411 BP; 397 A; 775 C; 659 G; 580 T; 0 other;

Query Match 93.3%; Score 14; DB 20; Length 2411;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTCTTGATCGGC 15
 |||||
 Db 157 CCTCTTGATCGGC 144

RESULT 10
 AAZ01425/c
 ID AAZ01425 standard; DNA; 1038602 BP.

XX AC AAZ01425;
 XX 07-OCT-1999 (first entry)
 XX Complete genome sequence of Chlamydia trachomatis.
 DE
 XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis; ss.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WO9928475-A2.
 XX
 PD 10-JUN-1999.
 XX
 PF 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 DR WPI; 1999-371125/31.
 XX
 XX Genome sequence of Chlamydia trachomatis

PS Claim 1; Page 373-656; 1755pp; English.
 XX
 XX The present sequence represents the complete genome of Chlamydia

CC trachomatis. Open reading frames (ORFs) of the genome encode
 CC polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines
 CC against Chlamydia trachomatis. Antisense and ribozyme sequences can also
 CC be used to control growth of the microorganism. Chlamydia trachomatis is
 CC responsible for a large number of diseases, e.g. eye diseases such as
 CC conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion
 CC conjunctivitis; genital diseases such as nongonococcal urethritis,
 CC epididymitis, cervicitis, salpingitis, perihepatitis, bartholinitis;
 CC pneumopathy in breast feeding infants; and venereal
 CC lymphogranulomatosis. The polypeptides of the invention may be of use in
 CC treating these diseases.
 XX
 SQ Sequence 1038602 BP; 304265 A; 214645 C; 214259 G; 305001 T; 432 other;

Query Match 93.3%; Score 14; DB 20; Length 1038602;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGG 14
 |||||
 Db 683576 ACCTCTTGATCGG 683563

RESULT 11
 AAS34099/c

ID AAS34099 standard; cDNA; 329 BP.
 XX
 AC AAS34099;
 XX 17-DEC-2001 (first entry)
 DT
 DE Human cDNA encoding a novel foetal antigen, SEQ ID NO 623.
 XX
 KW Human; foetal tissue antigen; ss; anti-inflammatory; neuroprotective;
 KW immunomodulator; cardiovascular; cytosolic; nephrotoxic;
 KW cardiovascular; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorder; breast neoplasm; cancer;
 KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
 KW cerebral ischaemia; angiogenesis; nervous system disorder;
 KW Alzheimer's disease; infection; ocular disorder; corneal infection;
 KW wound healing; epithelial cell proliferation; food additive.
 KW
 XX Homo sapiens.
 OS
 XX
 PN WO200155312-A2.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01321.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
DR WPI; 2001-488782/53.
XX P-PSDB; AAU21279.

PT New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems
XX
XX
PS Claim 1; SEQ ID No 623; 642pp; English.

XX The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence

Query Match 89.3%; Score 13.4; DB 22; Length 329;
Best Local Similarity 93.3%; Pred. No. 4.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTCTTGATCGGC 15
| | | | | | | | | | | | | | |
Db 121 ACCTCTTGATTTGGC 107

RESULT 12
AAQ66413/C
ID AAQ66413 standard; DNA; 360 BP.
XX
AC AAQ66413;
XX
DT 26-JAN-1995 (first entry)
XX
DE alca promoter.

XX Promoter; alca; crea binding site; glucose repression; ds.
 XX Aspergillus nidulans.
 OS
 XX WO9413820-A.
 XX PD 23-JUN-1994.
 XX PF 10-DEC-1993; 93WO-EP03553.
 XX PR 10-DEC-1992; 92US-0988778.
 XX PA (KONN) GIST-BROCADES NV.
 XX PI Hintz WE, Lagosky PA;
 XX DR WPI; 1994-217893/26.
 XX PT DNA constructs comprising a filamentous fungal promoter variant
 PT having a disrupted crea binding site - mediates expression of the
 PT protein in the presence of glucose
 XX
 PS Disclosure; Page 13; 21pp; English.
 XX
 CC A region of the A. nidulans alca promoter is given in AA066413. To
 CC construct an alca promoter variant, the crea binding sites were
 CC disrupted, but spatially conserved, by replacement of residues.
 CC The promoter variant allows expression of foreign genes in
 CC Aspergillus nidulans without the need to establish glucose-depleted
 CC growth conditions.
 CC
 SQ Sequence 360 BP; 97 A; 111 C; 83 G; 69 T; 0 other;
 QY 1 ACCTCTTGATCGGC 15
 DB 106 ACCTCTTGATCGGC 92
 Query Match 89.3%; Score 13.4; DB 15; Length 360;
 Best Local Similarity 93.3%; Pred. No. 4.4e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 13
 AAS03486
 ID AAS03486 standard; cDNA; 366 BP.
 XX
 AC AAS03486;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 73.
 XX
 KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 KW heart disease; complementarity determining region; CDR; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200127279-A1.
 XX
 PD 19-APR-2001.
 XX
 PF 11-OCT-2000; 2000WO-GB03900.
 XX
 PR 12-OCT-1999; 99US-0158812.
 XX
 PA (CAME-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 XX
 PI Edwards BM, Main SH, Vaughan TJ;
 XX WPI; 2001-282031/29.
 DR P-PSDB; AAU02586.
 XX

PT Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -
 XX
 PS Disclosure; Page 146; 182pp; English.
 XX
 CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.
 CC
 SQ Sequence 366 BP; 89 A; 97 C; 101 G; 79 T; 0 other;
 QY 1 ACCTCTTGATCGGC 15
 DB 91 ACCTATTTGATCGGC 105
 Query Match 89.3%; Score 13.4; DB 22; Length 366;
 Best Local Similarity 93.3%; Pred. No. 4.4e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 14
 AAH66453
 ID AAH66453 standard; DNA; 375 BP.
 XX
 AC AAH66453;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum coding sequence fragment SEQ ID NO: 1488.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX WPI; 2001-376931/40.
 DR P-PSDB; AAG91234.
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX
 PS Claim 8; SEQ ID NO: 1488; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 375 BP; 94 A; 96 C; 91 G; 94 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 375;
Best Local Similarity 93.3%; Pred. No. 4.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
| | | | | | | | | | | | | | | | | |
Db 60 ACCTCTTGATGGGC 74

RESULT 15
AAS03465
ID AAS03465 standard; cDNA; 387 BP.
XX
AC AAS03465;
XX
DT 29-AUG-2001 (first entry)
XX
DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 53.
XX
KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
KW heart disease; complementarity determining region; CDR; ss.
XX
OS Homo sapiens.
XX
PN WO200127279-A1.
XX
PD 19-APR-2001.
XX
PF 11-OCT-2000; 2000WO-GB03900.
XX
PR 12-OCT-1999; 99US-0158812.
XX
PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
XX
PI Edwards BM, Main SH, Vaughan TJ;
XX
DR WPI: 2001-282031/29.
DR P-PSDB; AAU02565.
XX
PT Panel of specific binding members of antibody molecules which bind to
PT whole adipocytes is used in the treatment of obesity and obesity
PT related diseases -
XX
PS Disclosure; Page 133; 182pp; English.
XX
CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
CC and light chain coding sequences of the invention. The antibodies can be
CC used in the treatment of obesity and obesity related diseases. The
CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
CC mass of an obese patient or the antibody can be used as a therapeutic
CC itself. Antibodies binding specifically to adipocytes can be used to
CC activate the immune system to destroy the cells by complement mediated
CC lysis. The antibodies may be labeled with a detectable label such as
CC radiolabel, fluorescent or chemical group and used in methods of
CC diagnosis in human subjects e.g. to determine the presence of adipocyte
CC antigen on the surface of an adipocyte to detect or determine the

CC presence or level of adipocytes in a cell or tissue sample. The
CC antibodies can be used as an alternative means of treatment for obese
CC patients other than undergoing surgery to remove excess fat. Antibodies
CC for different types of fat deposits can also be produced e.g. intra-
CC abdominal fat associated with heart disease.
XX
SQ Sequence 387 BP; 85 A; 107 C; 111 G; 84 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 387;
Best Local Similarity 93.3%; Pred. No. 4.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
| | | | | | | | | | | | | | | | | |
Db 91 AACTCTTGATCGGC 105

Search completed: June 26, 2003, 03:11:45
Job time : 88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-1
Perfect score: 15
Sequence: 1 acccttgcgcgc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	15	100.0	15	4	US-08-918-148-1
2	14	93.3	1443	3	US-08-959-381A-3
3	14	93.3	1626	3	US-08-959-381A-4
4	13.4	89.3	39	3	US-08-448-619-2
5	13.4	89.3	39	3	US-08-448-619-3
6	13.4	89.3	60	1	US-08-470-958-11
7	13.4	89.3	360	1	US-08-321-474-10
8	13.4	89.3	1177	2	US-08-841-349-11
9	13.4	89.3	4403765	4	US-09-103-840A-2
10	13.4	89.3	4411529	4	US-09-103-840A-1
11	13	86.7	9919	3	US-08-880-179-1
12	13	86.7	9936	4	US-08-972-927-2
13	12.4	82.7	213	1	US-08-247-907A-3
14	12.4	82.7	213	1	US-08-452-772-3
15	12.4	82.7	213	4	US-09-414-234-3
16	12.4	82.7	213	4	US-08-919-850-3
17	12.4	82.7	213	5	PCT-US94-05288-3
18	12.4	82.7	388	4	US-09-056-556-178
19	12.4	82.7	388	4	US-09-072-596-173
20	12.4	82.7	445	4	US-09-385-982-467
21	12.4	82.7	472	2	US-08-975-316-16
22	12.4	82.7	472	4	US-09-615-192A-16
23	12.4	82.7	567	4	US-09-615-192A-100
24	12.4	82.7	917	1	US-08-474-633A-6
25	12.4	82.7	917	4	US-08-823-771-6
26	12.4	82.7	917	5	PCT-US93-02480-6
27	12.4	82.7	985	4	US-09-056-556-182

C	28	12.4	82.7	985	4	US-09-072-596-177	Sequence 177, App
	29	12.4	82.7	1230	4	US-09-387-574-5	Sequence 5, Appli
	30	12.4	82.7	1230	4	US-09-668-096-5	Sequence 5, Appli
C	31	12.4	82.7	1270	1	US-08-247-907A-10	Sequence 10, Appl
C	32	12.4	82.7	1270	1	US-08-452-772-10	Sequence 10, Appl
C	33	12.4	82.7	1270	4	US-09-414-234-10	Sequence 10, Appl
C	34	12.4	82.7	1270	4	US-08-919-850-10	Sequence 10, Appl
C	35	12.4	82.7	1270	5	PCT-US94-05288-10	Sequence 10, Appl
C	36	12.4	82.7	1393	2	US-08-765-875-1	Sequence 1, Appli
C	37	12.4	82.7	1393	3	US-08-795-671-1	Sequence 1, Appli
C	38	12.4	82.7	1411	1	US-08-674-168-18	Sequence 18, Appl
C	39	12.4	82.7	1411	3	US-08-985-908-14	Sequence 14, Appl
C	40	12.4	82.7	1411	3	US-08-852-730-19	Sequence 19, Appl
	41	12.4	82.7	1455	4	US-09-615-192A-245	Sequence 245, App
	42	12.4	82.7	1670	4	US-09-092-315-4	Sequence 4, Appli
	43	12.4	82.7	1764	2	US-08-918-723-2	Sequence 2, Appli
	44	12.4	82.7	1764	2	US-09-237-507-2	Sequence 2, Appli
	45	12.4	82.7	1946	4	US-09-029-755C-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-918-148-1
; Sequence 1, Application US/08918148A
; Patent No. 6342220
; GENERAL INFORMATION:
; APPLICANT: Adams, Camellia
; APPLICANT: W.
; APPLICANT: Carter, Paul J.
; APPLICANT: Fendly, Brian M.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Agonist Antibodies
; FILE REFERENCE: P0979
; CURRENT APPLICATION NUMBER: US/08/918,148A
; CURRENT FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 79
; SEQ ID NO 1
; LENGTH: 15
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; NAME/KEY: 10F6scfv VH CDRI
; LOCATION: 1-15
; OTHER INFORMATION:
US-08-918-148-1

Query Match 100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCCTCTTGATCGGC 15
Db 1 ACCCTCTTGATCGGC 15

RESULT 2
US-08-959-381A-3/c
; Sequence 3, Application US/08959381A
; Patent No. 6048711
; GENERAL INFORMATION:
; APPLICANT: HINUMA, SHUJI
; APPLICANT: FUKUSUMI, SHOJI
; APPLICANT: KAWAMATA, YUJI
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA


```
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,381A
; FILING DATE: 28-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 286823/1996
; FILING DATE: 29-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
; REFERENCE/DOCKET NUMBER: TAK-50003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1443 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-959-381A-3

Query Match          93.3%; Score 14; DB 3; Length 1443;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 CCTCTTGATCGGC 15
        |||
Db      138 CCTCTTGATCGGC 125

RESULT 3
US-08-959-381A-4/c
; Sequence 4, Application US/08959381A
; Patent No. 6048711
; GENERAL INFORMATION:
; APPLICANT: HINUMA, SHUJI
; APPLICANT: FUKUSOMI, SHOJI
; APPLICANT: KANAMATA, YUJI
; TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR
; TITLE OF INVENTION: POLYNUCLEOTIDES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ratner & Prestia
; STREET: P.O. Box 980
; CITY: Valley Forge
; STATE: PA
; COUNTRY: USA
; ZIP: 19482
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/959,381A
; FILING DATE: 28-OCT-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 286823/1996
; FILING DATE: 29-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Prestia, Paul F
; REGISTRATION NUMBER: 23,031
```

```
; REFERENCE/DOCKET NUMBER: TAK-50003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0700
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1626 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-959-381A-4
```

```
Query Match          93.3%; Score 14; DB 3; Length 1626;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 CCTCTTGATCGGC 15
        |||
Db      321 CCTCTTGATCGGC 308
```

```
RESULT 4
US-08-448-619-2/c
; Sequence 2, Application US/08448619
; Patent No. 6140059
; GENERAL INFORMATION:
; APPLICANT: Schwallert, Manfred
; TITLE OF INVENTION: METHOD FOR THE OBTENTION OF NATIVE
; TITLE OF INVENTION: DOMAINS OF VIRAL MEMBRANE PROTEINS, THEIR USE, ESPECIALLY
; TITLE OF INVENTION: AS VACCINE AGAINST HIV, AND THESE NATIVE DOMAINS OF VIRAL
; TITLE OF INVENTION: MEMBRANE PROTEINS THEMSELVES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hardaway Law Firm
; STREET: P.O. Box 10107 Federal Station
; CITY: Greenville
; STATE: SC
; COUNTRY: USA
; ZIP: 29603-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/448,619
; FILING DATE: 29-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/DE94/00022
; FILING DATE: 12-JAN-1994
; APPLICATION DATA:
; APPLICATION NUMBER: DE P 43 01 017.2
; FILING DATE: 16-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hardaway III, John B.
; REGISTRATION NUMBER: 26,554
; REFERENCE/DOCKET NUMBER: RPE-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 864-233-6700
; TELEFAX: 864-233-2284
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 39 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Oligonucleotide"
; US-08-448-619-2
```


PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/988,778
; FILING DATE: 10-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mitchard, Leonard C.
; REGISTRATION NUMBER: 29009
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELETYPE: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-321-474-10

Query Match 89.3%; Score 13.4; DB 1; Length 360;
Best Local Similarity 93.3%; Pred. No. 76;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 106 ACCTCTTGACACGC 92

RESULT 8
US-08-841-349-11/c
; Sequence 11, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO4470050
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1177
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: For all n's in this sequence, n=(a or g or c or t)
; US-08-841-349-11

Query Match 89.3%; Score 13.4; DB 2; Length 1177;
Best Local Similarity 93.3%; Pred. No. 84;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 836 ACCTCTTGACCGGC 822

RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
; US-09-103-840A-2

Query Match 89.3%; Score 13.4; DB 4; Length 4403765;
Best Local Similarity 93.3%; Pred. No. 63;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 2279627 ACCTCTTGATCGAC 2279641

RESULT 10
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
; US-09-103-840A-1

Query Match 89.3%; Score 13.4; DB 4; Length 4411529;
Best Local Similarity 93.3%; Pred. No. 63;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 2277328 ACCTCTTGATCGAC 2277342

RESULT 11
US-08-880-179-1
; Sequence 1, Application US/08880179
; Patent No. 6091004
; GENERAL INFORMATION:
; APPLICANT: Ryals, John
; APPLICANT: Delaney, Terry
; APPLICANT: Friedrich, Leslie
; APPLICANT: Weymann, Kristiana
; APPLICANT: Lawton, Kay
; APPLICANT: Ellis, Daniel
; APPLICANT: Uknes, Scott
; APPLICANT: Jesse, Taco
; APPLICANT: Vos, Pieter
; TITLE OF INVENTION: GENE ENCODING A PROTEIN INVOLVED IN THE
; TITLE OF INVENTION: SIGNAL TRANSDUCTION CASCADE LEADING TO SYSTEMIC ACQUIRED RE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NO. 6091004artis Corporation
; STREET: 520 White Plains Road, P.O. Box 2005

```

; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/880,179
; FILING DATE:
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: CGC 1909
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9919 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-880-179-1

Query Match      86.7%; Score 13; DB 3; Length 9919;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCTCTTGATCG 13
Db      8951 ACCTCTTGATCG 8963

RESULT 12
; US-08-972-927-2/c
; Sequence 2, Application US/08972927
; Patent No. 6166290
; GENERAL INFORMATION:
; APPLICANT: Rea, Philip A
; APPLICANT: Lu, Yu-Ping
; APPLICANT: Li, Ze-Sheng
; TITLE OF INVENTION: GLUTATHIONE-S-CONJUGATE TRANSPORT IN
; TITLE OF INVENTION: PLANTS
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: One Commerce Square, 2005 Market Street, 22nd
; STREET: Floor
; CITY: Philadelphia
; STATE: Pennsylvania
; COUNTRY: US
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/972,927
; FILING DATE: 18-NOV-1997
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,040
; FILING DATE: 18-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/061,328
```

```

; FILING DATE: 08-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Doyle Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9596-1202
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9936 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-972-927-2

Query Match      86.7%; Score 13; DB 4; Length 9936;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCTCTTGATCG 13
Db      6598 ACCTCTTGATCG 6586

RESULT 13
; US-08-247-907A-3/c
; Sequence 3, Application US/08247907A
; Patent No. 5639638
; GENERAL INFORMATION:
; APPLICANT: WOZNEY, John
; APPLICANT: CELESTE, Anthony J.
; TITLE OF INVENTION: BMP-11 COMPOSITIONS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENETICS INSTITUTE, INC.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,907A
; FILING DATE: May 20, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LAZAR, Steven R.
; REGISTRATION NUMBER: 32,618
; REFERENCE/DOCKET NUMBER: G15205-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 876-1170
; TELEFAX: 617 876-5851
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; STRAIN: Human Activin WC
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..183
; FEATURE:
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-1
Perfect score: 15
Sequence: 1 acctcttgatcgcc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14	93.3	1446	9	US-10-083-168-17	Sequence 17, Appl
2	14	93.3	1446	9	US-10-083-168-86	Sequence 86, Appl
3	14	93.3	1446	9	US-10-251-385-81	Sequence 81, Appl
4	14	93.3	1446	9	US-10-251-385-207	Sequence 207, App
5	14	93.3	1728	10	US-09-826-508-1	Sequence 1, Appli
6	13.4	89.3	375	9	US-09-738-626-1488	Sequence 1488, Ap
7	13.4	89.3	390	10	US-09-867-701-1577	Sequence 1577, Ap
8	13.4	89.3	466	10	US-09-205-658-316	Sequence 316, App
9	13.4	89.3	593	10	US-09-998-598-1600	Sequence 1600, Ap
10	13.4	89.3	1350	9	US-09-738-626-252	Sequence 252, App
11	13.4	89.3	3239	12	US-10-044-090-438	Sequence 438, App
12	13.4	89.3	7263	9	US-10-037-182-19	Sequence 19, Appl
13	13.4	89.3	7554	9	US-10-037-182-17	Sequence 17, Appl
14	13.4	89.3	16489	9	US-09-764-868-1483	Sequence 1483, Ap
15	13.4	89.3	3309400	9	US-09-738-626-1	Sequence 1, Appli
16	13	86.7	2144	10	US-09-821-839-1	Sequence 1, Appli
17	13	86.7	3970	10	US-09-821-839-4	Sequence 4, Appli
18	13	86.7	9919	12	US-10-079-035-1	Sequence 1, Appli
19	12.4	82.7	25	9	US-10-098-263B-58767	Sequence 58767, A

20	12.4	82.7	25	9	US-10-098-263B-95238	Sequence 95238, A
21	12.4	82.7	186	10	US-09-783-590-2255	Sequence 2255, Ap
22	12.4	82.7	213	9	US-10-029-016-3	Sequence 3, Appli
23	12.4	82.7	300	10	US-09-294-093B-4484	Sequence 4484, Ap
24	12.4	82.7	306	10	US-09-294-093B-5109	Sequence 5109, Ap
25	12.4	82.7	307	9	US-10-015-219-175	Sequence 175, App
26	12.4	82.7	307	10	US-09-777-564-175	Sequence 175, App
27	12.4	82.7	316	10	US-09-864-761-29857	Sequence 29857, A
28	12.4	82.7	318	7	US-08-781-986A-4348	Sequence 4348, Ap
29	12.4	82.7	327	9	US-10-071-499A-10	Sequence 10, Appl
30	12.4	82.7	334	9	US-09-796-692-9041	Sequence 9041, Ap
31	12.4	82.7	334	9	US-10-040-862-9041	Sequence 9041, Ap
32	12.4	82.7	393	9	US-10-083-357-16	Sequence 16, Appl
33	12.4	82.7	396	9	US-09-970-966-167	Sequence 167, App
34	12.4	82.7	396	10	US-09-825-294-167	Sequence 167, App
35	12.4	82.7	401	10	US-09-998-598-8	Sequence 8, Appli
36	12.4	82.7	405	9	US-09-764-891-1978	Sequence 1978, Ap
37	12.4	82.7	409	10	US-09-960-352-927	Sequence 927, App
38	12.4	82.7	411	9	US-09-918-995-34920	Sequence 34920, A
39	12.4	82.7	417	9	US-09-918-995-7377	Sequence 7377, Ap
40	12.4	82.7	422	10	US-09-728-445-327	Sequence 327, App
41	12.4	82.7	445	9	US-09-871-161-467	Sequence 467, App
42	12.4	82.7	475	9	US-09-918-995-23638	Sequence 23638, A
43	12.4	82.7	487	9	US-09-918-995-26811	Sequence 26811, A
44	12.4	82.7	495	9	US-09-918-995-14852	Sequence 14852, A
45	12.4	82.7	498	10	US-09-954-456-588	Sequence 588, App

ALIGNMENTS

RESULT 1
US-10-083-168-17/c
; Sequence 17, Application US/10083168
; Publication No. US20030023069A1
; GENERAL INFORMATION:
; APPLICANT: Liaw, Chen W.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Behan, Dominic P.
; APPLICANT: Maciejewski-Lenior, Dominique
; APPLICANT: Leonard, James N.
; APPLICANT: Ortuno, Daniel
; APPLICANT: Lin, T-Lin
; TITLE OF INVENTION: Endogenous And No. US20030023069A1-Endogenous, Constitutively
; TITLE OF INVENTION: Receptors
; FILE REFERENCE: AREN-0320
; CURRENT APPLICATION NUMBER: US/10/083,168
; CURRENT FILING DATE: 2002-02-26
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 1446
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-083-168-17

Query Match 93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
Db 138 CCTCTTGATCGGC 125

RESULT 2
US-10-083-168-86/c
; Sequence 86, Application US/10083168
; Publication No. US20030023069A1
; GENERAL INFORMATION:
; APPLICANT: Liaw, Chen W.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Behan, Dominic P.

APPLICANT: Maciejewski-Lehior, Dominique
APPLICANT: Leonard, James N.
APPLICANT: Ortuno, Daniel
APPLICANT: Lin, I-Lin
TITLE OF INVENTION: Endogenous And No. US20030023069A1-Endogenous, Constitutively Act
TITLE OF INVENTION: Receptors
FILE REFERENCE: AREN-0320
CURRENT APPLICATION NUMBER: US/10/083,168
CURRENT FILING DATE: 2002-02-26
NUMBER OF SEQ ID NOS: 102
SOFTWARE: PatentIn version 3.1
SEQ ID NO 86
LENGTH: 1446
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: No. US20030023069A1el Sequence
US-10-083-168-86

Query Match 93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
|||||
Db 138 CCTCTTGATCGGC 125

RESULT 3
US-10-251-385-81/c
Sequence 81, Application US/10251385
Publication No. US20030105292A1
GENERAL INFORMATION:
APPLICANT: Behan, Dominic P.
APPLICANT: Chalmers, Derek T.
APPLICANT: Liaw, Chen W.
TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
TITLE OF INVENTION: Protein-Coupled
FILE REFERENCE: AREN-0040
CURRENT APPLICATION NUMBER: US/10/251,385
CURRENT FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: US/09/170,496
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 294
SOFTWARE: PatentIn version 3.1
SEQ ID NO 81
LENGTH: 1446
TYPE: DNA
ORGANISM: Homo sapiens
US-10-251-385-81

Query Match 93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
|||||
Db 138 CCTCTTGATCGGC 125

RESULT 4
US-10-251-385-207/c
Sequence 207, Application US/10251385
Publication No. US20030105292A1
GENERAL INFORMATION:
APPLICANT: Behan, Dominic P.
APPLICANT: Chalmers, Derek T.
APPLICANT: Liaw, Chen W.
TITLE OF INVENTION: No. US20030105292A1-Endogenous, Constitutively Activated Human G
TITLE OF INVENTION: Protein-Coupled
FILE REFERENCE: AREN-0040

CURRENT APPLICATION NUMBER: US/10/251,385
CURRENT FILING DATE: 2002-09-20
PRIOR APPLICATION NUMBER: US/09/170,496
PRIOR FILING DATE: 1998-10-13
NUMBER OF SEQ ID NOS: 294
SOFTWARE: PatentIn version 3.1
SEQ ID NO 207
LENGTH: 1446
TYPE: DNA
ORGANISM: Homo sapiens
US-10-251-385-207

Query Match 93.3%; Score 14; DB 9; Length 1446;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
|||||
Db 138 CCTCTTGATCGGC 125

RESULT 5
US-09-826-508-1/c
Sequence 1, Application US/09826508
Patent No. US20010025099A1
GENERAL INFORMATION:
APPLICANT: Nabil Elshourbagy
APPLICANT: Lisa Yawter
TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
FILE REFERENCE: GP-70744USB
CURRENT APPLICATION NUMBER: US/09/826,508
CURRENT FILING DATE: 2001-04-05
NUMBER OF SEQ ID NOS: 40
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 1728
TYPE: DNA
ORGANISM: HOMO SAPIENS
US-09-826-508-1

Query Match 93.3%; Score 14; DB 10; Length 1728;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTCTTGATCGGC 15
|||||
Db 287 CCTCTTGATCGGC 274

RESULT 6
US-09-738-626-1488
Sequence 1488, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07

```
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1488
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1488
```

```
Query Match          89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ACCTCTTGATCGGC 15
          |||
Db      60 ACCTCTTGATGGGC 74
```

RESULT 7

```
US-09-867-701-1577
; Sequence 1577, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1577
; LENGTH: 390
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(390)
; OTHER INFORMATION: n = A,T,C or G
US-09-867-701-1577
```

```
Query Match          89.3%; Score 13.4; DB 10; Length 390;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ACCTCTTGATCGGC 15
          |||
Db      84 ACCTCTTGACCGGC 98
```

RESULT 8

```
US-09-205-658-316/c
; Sequence 316, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 316
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-09-205-658-316
```

```
Query Match          89.3%; Score 13.4; DB 10; Length 466;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ACCTCTTGATCGGC 15
          |||
Db      258 ACCTCTTGATGGGC 244
```

RESULT 9

```
US-09-998-598-1600/c
; Sequence 1600, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2605
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 1600
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-1600
```

```
Query Match          89.3%; Score 13.4; DB 10; Length 593;
Best Local Similarity 93.3%; Pred. No. 2.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 ACCTCTTGATCGGC 15
          |||
Db      373 ACCTCTTGATGGGC 359
```

RESULT 10

```
US-09-738-626-252/c
; Sequence 252, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
```


SEQ ID NO 252
LENGTH: 1350
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-252

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 9; Length 1350;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 113 ACCTCTTGATCGGC 99

RESULT 11
US-10-044-090-438

Sequence 438, Application US/10044090
Patent No. US20020137081A1
GENERAL INFORMATION:

APPLICANT: Olga Bandman
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
FILE REFERENCE: PA-0028 US
CURRENT APPLICATION NUMBER: US/10/044,090
PRIOR FILING DATE: 2002-01-09
NUMBER OF SEQ ID NOS: 850
SOFTWARE: PERL Program
SEQ ID NO 438
LENGTH: 3239
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: Incyte ID No. US20020137081A1 241624.1
US-10-044-090-438

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 12; Length 3239;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 1958 ACCTCTTGATCGGC 1972

RESULT 12
US-10-037-182-19

Sequence 19, Application US/10037182
Publication No. US20030044899A1
GENERAL INFORMATION:

APPLICANT: Tryggyvason, Karl
APPLICANT: Doi, Masayuki
APPLICANT: Thyboll, Jill
TITLE OF INVENTION: Recombinant Laminin 10
FILE REFERENCE: 99-274-F
CURRENT APPLICATION NUMBER: US/10/037,182
PRIOR FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,449
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/279,282
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 19
LENGTH: 7263
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(4716)
US-10-037-182-19

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 9; Length 7263;

Best Local Similarity 93.3%; Pred. No. 2.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 4445 ACCTCTTGATCGGC 4459

RESULT 13
US-10-037-182-17

Sequence 17, Application US/10037182
Publication No. US20030044899A1
GENERAL INFORMATION:
APPLICANT: Tryggyvason, Karl
APPLICANT: Doi, Masayuki
APPLICANT: Thyboll, Jill
TITLE OF INVENTION: Recombinant Laminin 10
FILE REFERENCE: 99-274-F
CURRENT APPLICATION NUMBER: US/10/037,182
PRIOR FILING DATE: 2001-12-21
PRIOR APPLICATION NUMBER: 60/257,449
PRIOR FILING DATE: 2000-12-21
PRIOR APPLICATION NUMBER: 60/279,282
PRIOR FILING DATE: 2001-03-28
NUMBER OF SEQ ID NOS: 36
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 17
LENGTH: 7554
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (193)..(5007)
NAME/KEY: sig_peptide
LOCATION: (193)..(291)
US-10-037-182-17

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 9; Length 7554;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 4736 ACCTCTTGATCGGC 4750

RESULT 14
US-09-764-868-1483

Sequence 1483, Application US/09764868
Patent No. US20020168711A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT232
CURRENT APPLICATION NUMBER: US/09/764,868
PRIOR FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 1510
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1483
LENGTH: 16489
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-868-1483

Query Match
Best Local Similarity 89.3%; Score 13.4; DB 9; Length 16489;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
Db 16177 ACCTCTTGATCGGC 16191

RESULT 15

US-09-738-626-1/c

; Sequence 1, Application US/09738626
; Publication No. US20020197605A1

GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 89.3%; Score 13.4; DB 9; Length 3309400;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 238288 ACCTCTTGATCGGC 238274

Search completed: June 26, 2003, 04:56:43
Job time : 75.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-1
Perfect score: 15
Sequence: 1 acctcttgatcgac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	15	100.0	237	14	BQ757051	BQ757051 EBem10_SQ
C 2	15	100.0	433	13	BM376561	BM376561 EBem05_SQ
C 3	15	100.0	441	9	AJ434075	AJ434075 AJ434075
C 4	15	100.0	444	14	BQ461961	BQ461961 HD02L15r
C 5	15	100.0	458	13	BI777123	BI777123 EBro03_SQ
C 6	15	100.0	474	14	BQ468011	BQ468011 HR01K11r

C	7	15	100.0	518	12	BG300356	BG300356 HVSMEB001
C 8	15	100.0	521	13	BI780309	BI780309 EBem06_SQ	
C 9	15	100.0	558	10	AV941160	AV941160 AV941160	
C 10	15	100.0	565	14	BQ465771	BQ465771 HU04K09r	
C 11	15	100.0	575	14	BQ471255	BQ471255 HV01N14T	
C 12	15	100.0	597	14	BQ464764	BQ464764 HU01G04T	
C 13	15	100.0	600	10	AV915369	AV915369 AV915369	
C 14	15	100.0	607	17	BH474182	BH474182 BOGVMB8TF	
C 15	15	100.0	607	14	BQ465677	BQ465677 HU04F22r	
C 16	15	100.0	615	14	BQ471432	BQ471432 HV02G15r	
C 17	15	100.0	620	13	BM099080	BM099080 EBP105_SQ	
C 18	15	100.0	636	12	BG415523	BG415523 HVSMEK000	
C 19	15	100.0	653	12	BG415434	BG415434 HVSMEK000	
C 20	15	100.0	670	17	BH446702	BH446702 BOGVJ04TR	
C 21	15	100.0	671	13	BI954106	BI954106 HVSMEK001	
C 22	15	100.0	799	12	BF254300	BF254300 HVSMEK002	
C 23	15	100.0	851	12	BG418413	BG418413 HVSMEK002	
C 24	15	100.0	932	12	BF623261	BF623261 HVSMEK001	
C 25	15	100.0	1037	13	BI950768	BI950768 HVSMEI002	
C 26	14	93.3	273	10	BE415681	BE415681 MML038.BO	
C 27	14	93.3	281	9	AL829578	AL829578 AL829578	
C 28	14	93.3	410	12	BF146084	BF146084 WHE1837.A	
C 29	14	93.3	475	13	BM027631	BM027631 CSAC30019	
C 30	14	93.3	499	13	BI805258	BI805258 S034E11.S	
C 31	14	93.3	512	9	AI773000	AI773000 EST254100	
C 32	14	93.3	524	14	BQ273893	BQ273893 rc72c10.Y	
C 33	14	93.3	537	13	BI669564	BI669564 603294282	
C 34	14	93.3	553	14	BQ853560	BQ853560 OGB20N12.	
C 35	14	93.3	568	13	BJ302594	BJ302594 BJ302594	
C 36	14	93.3	592	13	BJ260486	BJ260486 BJ260486	
C 37	14	93.3	607	12	BG910143	BG910143 602805504	
C 38	14	93.3	620	17	AZ432572	AZ432572 1M0218K01	
C 39	14	93.3	621	12	BF530127	BF530127 602040306	
C 40	14	93.3	624	13	BJ033194	BJ033194 BJ033194	
C 41	14	93.3	628	13	BI604112	BI604112 603244048	
C 42	14	93.3	632	13	BJ037202	BJ037202 BJ037202	
C 43	14	93.3	639	13	BJ034888	BJ034888 BJ034888	
C 44	14	93.3	640	10	BE427382	BE427382 PSR6374-B	
C 45	14	93.3	668	13	BI601027	BI601027 603249503	

ALIGNMENTS

RESULT 1
BQ757051/c 237 bp mRNA linear EST 26-JUL-2002
LOCUS
DEFINITION
BQ757051 Hordeum vulgare cDNA clone EBem10_SQ002_K07 5', mRNA
sequence.

ACCESSION
BQ757051
VERSION
BQ757051.1 GI:21965523

KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.

REFERENCE
1 (bases 1 to 237)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
Development of Barley Transcriptome Resources
Unpublished (2001)
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk.

FEATURES
source
1..237
/organism="Hordeum vulgare"

	/cultivar="Optic"	
	/db_xref="taxon:4513"	
	/clone="EBem10_SQ002_K07"	
	/clone_lib="embryo, 2 day germination, no treatment, cv	
	Optic, EBem10"	
	/tissue_type="embryo"	
	/dev_stage="2 Day germination"	
	/lab_host="DH10B"	
	/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;	
	Non-normalised library, directionally cloned into pSPORT1.	
	Derived from embryos dissected from germinating grains (2	
	day) in glasshouse grown barley plants. Developed as part	
	of the barley transcriptome resources of BBSRC/SERAD	
	funded cereal IGF (Investigating Gene Function) project."	
BASE COUNT	45 a 68 c 73 g 51 t	
ORIGIN		
Query Match	100.0%; Score 15; DB 14; Length 237;	
Best Local Similarity	100.0%; Pred. No. 6.1e+02;	
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 ACCTCTTGATCGGC 15 	
Dn	126 ACCCTTGGATCGGC 112	
RESULT 2		
BM376561/c	433 bp mRNA linear EST 23-JUL-2002	
LOCUS	EBem05_SQ002_J05_R embryo, 14 DPA, no treatment, cv Optic, EBem05	
DEFINITION	Hordeum vulgare cDNA clone EBem05_SQ002_J05 5', mRNA sequence.	
ACCESSION	BM376561	
VERSION	BM376561.2 GI:21934181	
KEYWORDS	EST.	
SOURCE	Hordeum vulgare.	
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum. 1 (bases 1 to 433) Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L., Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R. Development of Barley Transcriptome Resources Unpublished (2001)	
TITLE	On Jan 10, 2002 this sequence version replaced gi:18119951.	
JOURNAL	Contact: Maugh R, Marshall DF	
COMMENT	Genome Dynamics/Computational Biology Scottish Crop Research Institute Invergowrie, Dundee, DD2 5DA, Scotland, UK Tel: 00 44 1382 562731 Fax: 00 44 1382 562426 Email: est@scri.sari.ac.uk All sequence has a phred quality score of 20 or over Seq primer: M13 reverse. location/Qualifiers 1 . 433 organism="Hordeum vulgare" cultivar="Optic" db_xref="taxon:4513" clone="EBem05_SQ002_J05" clone_lib="embryo, 14 DPA, no treatment, cv Optic, EBem05" tissue_type="embryo" dev_stage="14 DPA" lab_host="DH10B" note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I; Non-normalised library, directionally cloned into pSPORT1. Derived from embryos dissected from developing grains (14 days post anthesis) in glasshouse grown barley plants. Developed as part of the barley transcriptome resources of BBSRC/SERAD funded cereal IGF (Investigating Gene Function) project." BASE COUNT 106 a 103 c 115 g 109 t	

[illegible]

Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 444 Std Error: 0.00
Plate: 2 row: 1 column: 15
Seq primer: M13rev.

FEATURES
source
1. .444
Location/Qualifiers

/organism="Hordeum vulgare"
/cultivar="Golden Promise"
/db_xref="taxon:4513"
/clone="HD02L15"
/tissue_type="HD"
/tissue_type="callus"
/dev_stage="callus (5-10 mm in diameter)"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1 kb"

BASE COUNT 105 a 114 c 121 g 104 t
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 444;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 153 ACCTCTTGATCGGC 139

RESULT 5
BI777123/c 458 bp mRNA linear EST 23-JUL-2002
LOCUS EBro03_SQ002_A18_R root, 3 week, waterlogged, cv Optic, EBro03
DEFINITION Hordeum vulgare cDNA clone EBro03_SQ002_A18 5', mRNA sequence.
ACCESSION BI777123
VERSION BI777123 GI:21948259
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare

REFERENCE 1 (bases 1 to 458)
AUTHORS Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L., Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
TITLE Development of Barley Transcriptome Resources
JOURNAL Unpublished (2001)
COMMENT On Sep 26, 2001 this sequence version replaced gi:15780015.
Contact: Waugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.

FEATURES
source
1. .458
Location/Qualifiers

/organism="Hordeum vulgare"
/cultivar="Optic"
/db_xref="taxon:4513"
/clone="EBro03_SQ002_A18"
/clone_lib="root, 3 week, waterlogged, cv Optic, EBro03"

/tissue_type="root"
/dev_stage="3 week"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
Non-normalised library, directionally cloned into pSPORT1.
Derived from roots of 3 week old waterlogged barley.
plants. Developed as part of the barley transcriptome
resources of BBSRC/SEERAD funded cereal IGF (Investigating
Gene Function) project."

BASE COUNT 111 a 116 c 121 g 110 t
ORIGIN

Query Match 100.0%; Score 15; DB 13; Length 458;
Best Local Similarity 100.0%; Pred. No. 7.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 153 ACCTCTTGATCGGC 139

RESULT 6
BQ468011/c 474 bp mRNA linear EST 30-MAY-2002
LOCUS HR01K11c HR Hordeum vulgare cDNA clone HR01K11 5-PRIME, mRNA
DEFINITION sequence.
ACCESSION BQ468011 GI:21275793
VERSION BQ468011.1
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare

REFERENCE 1 (bases 1 to 474)
AUTHORS Zhang, H., Weschke, W., Michalek, W., Stein, N. and Graner, A.
TITLE EST sequencing and analysis in barley (2002)
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 474 Std Error: 0.00
Plate: 1 row: K column: 11
Seq primer: M13rev.

FEATURES
source
1. .474
Location/Qualifiers

/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HR01K11"
/clone_lib="HR"
/tissue_type="roots"
/dev_stage="grown on filter paper at 25 Grad C for 48
hours in petridish, 1-1.5 cm in length"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 800 bp"

BASE COUNT 114 a 120 c 130 g 110 t
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 474;
Best Local Similarity 100.0%; Pred. No. 7.6e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACCTCTTGATCGGC 15
 |||||||||||||
 Db 153 ACCTCTTGATCGGC 139

RESULT 7

BG300356/c

LOCUS 518 bp mRNA linear EST 22-OCT-2001
 DEFINITION HVSMEB0012021f Hordeum vulgare seedling shoot EST library
 HVCDNA0002 (Dehydration stress) Hordeum vulgare cDNA clone
 HVSMEB0012021f, mRNA sequence.

ACCESSION BG300356
 VERSION BG300356
 KEYWORDS GI:13097883
 SOURCE EST.

ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 518)
 Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
 , Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Choi, D.W., Fenton
 , R.D., Oates, R. and Main, D.
 Development of a genetically and physically anchored EST resource
 for barley genomics: Morex drought-stressed seedling shoot cDNA
 library

TITLE Unpublished (2001)

JOURNAL Contact: Wing RA
 COMMENT Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu

FEATURES
 source Total hg bases = 235
 Seq primer: AATTACCCCTCACTAAGG
 High quality sequence stop: 411.
 Location/Qualifiers
 1..518
 /organism="Hordeum vulgare"
 /cultivar="Morex"
 /db_xref="taxon:4513"
 /clone="HVSMEB0012021f"
 /clone_lib="Hordeum vulgare seedling shoot EST library
 HVCDNA0002 (Dehydration stress)"
 /tissue_type="Seedling shoot"
 /lab_host="TJ121"
 /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
 Seeds were surface sterilized then germinated under axenic
 conditions in the dark at room temperature on filter paper
 with water, nystatin and cefotaxime in covered
 crystallization dishes. Five-day old seedlings were
 incubated at 90% RH for 24 hr. Shoots were then harvested,
 total RNA was prepared, poly(A) RNA was purified, one
 primary unamplified cDNA library was made, 600000 pfu were
 in vivo excised to give pBluescript SK(-) cDNA phagemids.
 These steps were performed in the TJ Close laboratory at
 the University of California, Riverside (Choi, Close,
 Fenton). Phagemids were plated and picked at the Clemson
 University Genomics Institute (CUGI) (Begum, Palmer,
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
 sequencing and sequence analysis were performed at CUGI
 (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main).
 The sequence has been trimmed to remove vector sequence
 and contains a minimum of 100 bases of phred value 20 or
 above. For more details on library preparation and
 sequence analysis see
 http://www.genome.clemson.edu/projects/barley. To order
 this clone see http://www.genome.clemson.edu/orders Also
 see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
 Genetically and physically anchored EST resources for

barley genomics. Barley Genetics Newsletter 31:29-30.
 (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
 BASE COUNT 135 a 145 c 120 g 117 t 1 others
 ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 518;
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
 |||||||||||||
 Db 207 ACCTCTTGATCGGC 193

RESULT 8

BI780309/c

LOCUS 521 bp mRNA linear EST 23-JUL-2002
 DEFINITION EBem06_SQ001_E02_R embryo, 21 DPA, no treatment, cv Optic, EBem06
 Hordeum vulgare cDNA clone EBem06_SQ001_E02 5', mRNA sequence.

ACCESSION BI780309
 VERSION BI780309
 KEYWORDS GI:21934849
 SOURCE EST.

ORGANISM Hordeum vulgare.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
 ; Triticeae; Hordeum.

REFERENCE 1 (bases 1 to 521)
 Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardie, L.,
 Ramsay, L., Machray, G., Marshall, D.F.M. and Waugh, R.
 Development of Barley Transcriptome Resources
 Unpublished (2001)

TITLE On Sep 26, 2001 this sequence version replaced gi:15783201.

JOURNAL Contact: Waugh R, Marshall DP
 COMMENT Genome Dynamics/Computational Biology
 Scottish Crop Research Institute
 Invergowrie, Dundee, DD2 5DA, Scotland, UK
 Tel: 00 44 1382 562731
 Fax: 00 44 1382 562426
 Email: est@scri.sari.ac.uk
 All sequence has a Phred quality score of 20 or over
 Seq primer: M13 reverse.
 Location/Qualifiers
 1..521
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="EBem06_SQ001_E02"
 /clone_lib="embryo, 21 DPA, no treatment, cv Optic,
 EBem06"
 /tissue_type="embryo"
 /dev_stage="21 DPA"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from embryos dissected from developing grains (21
 days post anthesis) in glasshouse grown barley plants.
 Developed as part of the barley transcriptome resources of
 BBSRC/SEERAD funded cereal IGF (Investigating Gene
 Function) project."

FEATURES
 source Location/Qualifiers
 1..521
 /organism="Hordeum vulgare"
 /cultivar="Optic"
 /db_xref="taxon:4513"
 /clone="EBem06_SQ001_E02"
 /clone_lib="embryo, 21 DPA, no treatment, cv Optic,
 EBem06"
 /tissue_type="embryo"
 /dev_stage="21 DPA"
 /lab_host="DH10B"
 /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
 Non-normalised library, directionally cloned into pSPORT1.
 Derived from embryos dissected from developing grains (21
 days post anthesis) in glasshouse grown barley plants.
 Developed as part of the barley transcriptome resources of
 BBSRC/SEERAD funded cereal IGF (Investigating Gene
 Function) project."

BASE COUNT 127 a 121 c 145 g 128 t
 ORIGIN

Query Match 100.0%; Score 15; DB 13; Length 521;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
 |||||||||||||
 Db 136 ACCTCTTGATCGGC 122

RESULT 9

AV941160/c 558 bp mRNA linear EST 18-JAN-2002
LOCUS AV941160 K. Sato unpublished cDNA library, strain H602 adult,
DEFINITION heading stage top three leaves Hordeum vulgare subsp. spontaneum
cDNA clone bah31m09 5', mRNA sequence.
ACCESSION AV941160
VERSION AV941160.1 GI:18236957
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. spontaneum.
ORGANISM Hordeum vulgare subsp. spontaneum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 558)
AUTHORS Sato,K., Saisho,D. and Takeda,K.
TITLE Barley EST sequencing project in NIG and Okayama Univ
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-I
Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.
Location/Qualifiers
1..558
/organism="Hordeum vulgare subsp. spontaneum"
/strain="H602"
/db_xref="taxon:77009"
/clone="bah31m09"
/clone_lib="K. Sato unpublished cDNA library, strain H602
adult, heading stage top three leaves"
/tissue_type="top three leaves"
/dev_stage="adult, heading stage"
BASE COUNT 138 a 142 c 149 g 129 t
ORIGIN
Query Match 100.0%; Score 15; DB 10; Length 558;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTCTTGATCGGC 15
|||||
Db 169 ACCTCTTGATCGGC 155
RESULT 10 565 bp mRNA linear EST 30-MAY-2002
BQ465771/c
LOCUS BQ465771 HU04K09r HU Hordeum vulgare cDNA clone HU04K09 5-PRIME, mRNA
DEFINITION sequence.
ACCESSION BQ465771
VERSION BQ465771.1 GI:21273553
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 565)
AUTHORS Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
A.
TITLE Barley ESTs from germinating seeds
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 565 Std Error: 0.00
Plate: 4 row: K column: 9

Seq primer: M13rev.
Location/Qualifiers
1..565
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HU04K09"
/clone_lib="HU"
/tissue_type="germinating seeds"
/dev_stage="germinating seeds, 16-48 h"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds,
16-48h. Due to a cloning artefact caused by the kit, in
most cases the EcoRI site is NOT present, as well as the
EcoRI adapter used for cloning. To excise the insert,
restriction sites upstream EcoRI should be used (e.g.
BamHI, SalI, PstI). NOTE: Also due to the cloning system
used Blue/White selection for recombinants is not 100%
reliable. Average insert size is 1.2 Kb"
BASE COUNT 136 a 134 c 158 g 137 t
ORIGIN
Query Match 100.0%; Score 15; DB 14; Length 565;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTCTTGATCGGC 15
|||||
Db 148 ACCTCTTGATCGGC 134
RESULT 11 575 bp mRNA linear EST 30-MAY-2002
BQ471255/c
LOCUS BQ471255 HV01N14r HV Hordeum vulgare cDNA clone HV01N14 5-PRIME, mRNA
DEFINITION sequence.
ACCESSION BQ471255
VERSION BQ471255.1 GI:21279037
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 575)
AUTHORS Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner
A.
TITLE Barley ESTs from germinating seeds
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 575 Std Error: 0.00
Plate: 1 row: N column: 14
Seq primer: T3.
Location/Qualifiers
1..575
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HV01N14"
/clone_lib="HV"
/tissue_type="germinating seeds"
/dev_stage="germinating seeds (48-96 h)"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); Roots were grown for
two days on filter paper at room temperature. Due to a

cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI).
NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1 kb"

BASE COUNT 140 a 138 c 159 g 138 t
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 575;
Best Local Similarity 100.0%; Pred. No. 8.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 150 ACCTCTTGATCGGC 136

RESULT 12
BQ464764/c 597 bp mRNA linear EST 30-MAY-2002
LOCUS HU01G04T HU Hordeum vulgare cDNA clone HU01G04 5-PRIME, mRNA
DEFINITION sequence.

ACCESSION BQ464764
VERSION BQ464764.1 GI:21272546
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
AUTHORS Zhang, H., Potokina, E., Michalek, W., Weschke, W., Stein, N. and Graner, A.
TITLE Barley ESTs from germinating seeds
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 597 Std Error: 0.00
Plate: 1 row: 6 column: 4
Seq primer: T3.

FEATURES
source location/Qualifiers
1..597

/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HU01G04"
/clone_lib="HU"
/tissue_type="germinating seeds"
/dev_stage="germinating seeds, 16-48 h"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds, 16-48h. Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 1.2 kb"

BASE COUNT 149 a 148 c 160 g 140 t
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 597;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 165 ACCTCTTGATCGGC 151

RESULT 13
AV915369/c 600 bp mRNA linear EST 18-JAN-2002
LOCUS AV915369 K. Sato unpublished cDNA library, cv. Haruna Nijo
DEFINITION germination shoots Hordeum vulgare subsp. vulgare cDNA clone
bags14b14 5', mRNA sequence.

ACCESSION AV915369
VERSION AV915369
KEYWORDS AV915369.1 GI:18211146
SOURCE EST.
ORGANISM Hordeum vulgare subsp. vulgare.
Hordeum vulgare subsp. vulgare.

REFERENCE Sato, K., Saisho, D. and Takeda, K.
AUTHORS Barley EST sequencing project in NIG and Okayama Univ
TITLE Unpublished (2002)
JOURNAL Contact: Tadasu Shin-i
COMMENT Center for Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES
source location/Qualifiers
1..600

/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Haruna Nijo"
/db_xref="taxon:112509"
/clone="bags14b14"
/clone_lib="K. Sato unpublished cDNA library, cv. Haruna Nijo germination shoots"
/tissue_type="shoots"
/dev_stage="germination"
BASE COUNT 149 a 147 c 160 g 142 t 2 others
ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 600;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 163 ACCTCTTGATCGGC 149

RESULT 14
BH474182/c 606 bp DNA linear GSS 13-DEC-2001
LOCUS BGV888T BGV Brassica oleracea genomic clone BGV888, DNA
DEFINITION sequence.
ACCESSION BH474182
VERSION BH474182.1 GI:17682293
KEYWORDS GSS.
SOURCE Brassica oleracea.
ORGANISM Brassica oleracea.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
AUTHORS 1 (bases 1 to 606)
TITLE Town, C.D., Van Aken, S., Uterback, T. and Fraser, C.M.
JOURNAL Whole genome shotgun sequencing of Brassica oleracea
COMMENT Unpublished (2001)
Other_GSSs: BGV888T
CONTACT: Chris Town
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA.

Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@ligr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES
source
location/Qualifiers

1.606
/organism="Brassica oleracea"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOGYM88"
/clone_lib="BOGY"
/note="Vector: PHOS1; Site_1: BstXI; 2-3 kb sheared
genomic DNA inserted into PHOS1 using BstXI linkers"

BASE COUNT 150 a 148 c 128 g 180 t
ORIGIN

Query Match 100.0%; Score 15; DB 17; Length 606;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 330 ACCTCTTGATCGGC 316

RESULT 15
BQ465677/c 607 bp mRNA linear EST 30-MAY-2002
LOCUS HU04F22r HU Hordeum vulgare cDNA clone HU04F22 5-PRIME, mRNA

DEFINITION
sequence.
ACCESSION BQ465677
VERSION BQ465677.1 GI:21273459

KEYWORDS
SOURCE EST.
ORGANISM Hordeum vulgare.
Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
; Triticeae; Hordeum
1 (bases 1 to 607)
Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner

REFERENCE
AUTHORS 'A.
Barley ESTs from germinating seeds

TITLE
JOURNAL Unpublished (2002)
COMMENT Contact: Stein Nils

Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 607 Std Error: 0.00
Plate: 4 row: F column: 22
Seq primer: M13rev.

FEATURES
source
location/Qualifiers

1.607
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HU04F22"
/clone_lib="HU"
/tissue_type="germinating seeds"
/dev_stage="germinating seeds, 16-48 h"
/lab_host="XL10-Gold"
/note="Vector: pBluescript SK+; Site_1: EcoRI (5'-end of
cDNA); Site_2: XhoI (3'-end of cDNA); germinating seeds,
16-48h. Due to a cloning artefact caused by the kit, in
most cases the EcoRI site is NOT present, as well as the
EcoRI adapter used for cloning. To excise the insert,
restriction sites upstream EcoRI should be used (e.g.
BamHI, SalI, PstI). NOTE: Also due to the cloning system
used Blue/white selection for recombinats is not 100%

BASE COUNT 153 a 154 c 161 g 139 t
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 607;
Best Local Similarity 100.0%; Pred. No. 8.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTCTTGATCGGC 15
|||||
Db 179 ACCTCTTGATCGGC 165

Search completed: June 26, 2003, 04:50:05
Job time : 754.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	6 AR183480	AR183480 Sequence
2	15	100.0	165	9 HSTIMHCY10	Z29985 H.sapiens D
3	15	100.0	217	9 HSJ227738	AJ227738 Homo sapi
4	15	100.0	223	9 AF017465	AF017465 Homo sapi
5	15	100.0	223	9 HSJ227734	AJ227734 Homo sapi
6	15	100.0	246	9 HSA306240	AJ306240 Homo sapi
7	15	100.0	255	9 AF297164	AF297164 Homo sapi
8	15	100.0	258	9 HSEM112V3	Y12032 H.sapiens r
9	15	100.0	261	9 HSA406644	AJ406644 Homo sapi
10	15	100.0	273	9 HSA415895	AJ415895 Homo sapi
11	15	100.0	278	9 HSZ77321	Z77321 H.sapiens g
12	15	100.0	279	9 AF077461	AF077461 Homo sapi
13	15	100.0	283	9 AY003824	AY003824 Homo sapi
14	15	100.0	291	9 HUMIGL2E5	L04325 Human (cell
15	15	100.0	291	9 HSZ80382	Z80382 H.sapiens B
16	15	100.0	293	9 HSIG14	X67073 H.sapiens I
17	15	100.0	294	9 AF119800	AF119800 Homo sapi
18	15	100.0	294	9 AF119801	AF119801 Homo sapi
19	15	100.0	294	9 AF119802	AF119802 Homo sapi
20	15	100.0	294	9 AF119803	AF119803 Homo sapi
21	15	100.0	294	9 AF119804	AF119804 Homo sapi
22	15	100.0	294	9 AF151702	AF151702 Homo sapi
23	15	100.0	294	9 AF151704	AF151704 Homo sapi
24	15	100.0	294	9 HSU09062	U09062 Human immu
25	15	100.0	294	9 HSI09062	X92287 H.sapiens g
26	15	100.0	296	9 HSI09062	AX470326 Sequence
27	15	100.0	297	9 HSI09062	L03824 Human IG re
28	15	100.0	297	9 HSI09062	Z80419 H.sapiens B
29	15	100.0	297	9 HSI09062	Z12336 H.sapiens g
30	15	100.0	297	9 HSI09062	L03825 Human IG re
31	15	100.0	300	9 HSI09062	Z80625 H.sapiens B
32	15	100.0	300	9 HSI09062	AF077475 Homo sapi
33	15	100.0	300	9 HSI09062	AF077481 Homo sapi
34	15	100.0	300	9 HSI09062	AF077428 Homo sapi
35	15	100.0	304	9 AF077481	AF077481 Homo sapi
36	15	100.0	305	9 AF077428	AF077428 Homo sapi
37	15	100.0	306	9 AF077456	AF077456 Homo sapi
38	15	100.0	306	9 HSB83H08N	X87074 H.sapiens B
39	15	100.0	307	9 AF077478	AF077478 Homo sapi
40	15	100.0	309	9 AF077466	AF077466 Homo sapi
41	15	100.0	309	9 HSB83G07N	X87070 H.sapiens B
42	15	100.0	311	9 HSA289464	AJ289464 Homo sapi
43	15	100.0	312	9 AF077455	AF077455 Homo sapi
44	15	100.0	312	9 AF077471	AF077471 Homo sapi
45	15	100.0	312	9 AF077471	AF077471 Homo sapi

ALIGNMENTS

RESULT 1
AR183480
LOCUS AR183480 15 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 13 from patent US 6342220.
ACCESSION AR183480
VERSION AR183480.1 GI:20227449
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 13 29-JAN-2002;
FEATURES Location/Qualifiers

source 1..15
/organism="unknown"
BASE COUNT 5 a 4 c 3 g 3 t
ORIGIN

Query Match 100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
1 GACTACTACATGAGC 15

RESULT 2
HSIMHCV10
LOCUS
DEFINITION H.sapiens DNA for immunoglobulin heavy chain variable region (DA-10).
ACCESSION Z29985
VERSION Z29985.1 GI:505455
KEYWORDS immunoglobulin heavy chain variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165)
Cook,G.P., Tomlinson,I.M., Walter,G., Carter,N.P., Riethman,H., Bulwela,L., Winter and Rabbits,T.H.
Unpublished
JOURNAL 2 (bases 1 to 165)
REFERENCE Tomlinson,M.
AUTHORS Direct Submission
TITLE Submitted (10-FEB-1994) Ian Tomlinson, MRC Centre for Protein Engineering, Hills Road, Cambridge, CB2 2QH, U.K.
JOURNAL 3 (bases 1 to 165)
REFERENCE Cook,G.P., Tomlinson,I.M., Walter,G., Riethman,H., Carter,N.P., Bulwela,L., Winter,G. and Rabbits,T.H.
AUTHORS A map of the human immunoglobulin VH locus completed by analysis of the telomeric region of chromosome 14q
JOURNAL Nat. Genet. 7 (2), 162-168 (1994)
MEDLINE 95004581
PUBMED 7920635

FEATURES
source location/Qualifiers
1..165
/organism="Homo sapiens"
/isolate="der 8 chromosome from 'DAUDI' cell line"
/db_xref="taxon:9606"
/dev_stage="adult"
/germline
1..165
/pseudo
/codon_start=1
/product="immunoglobulin heavy chain variable region (DA-10)"
/db_xref="PID:e72321"
BASE COUNT 51 a 36 c 44 g 34 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 165;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
10 GACTACTACATGAGC 24

RESULT 3
HSJ227738
LOCUS
DEFINITION Homo sapiens DNA for rearranged immunoglobulin heavy chain gene, 4-1VH3.

ACCESSION AJ227738
VERSION AJ227738.1 GI:2951653
KEYWORDS heavy chain; immunoglobulin; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 217)
Delabie,J.
AUTHORS Direct Submission
JOURNAL Submitted (03-MAR-1998) Delabie J., Pathology, University of Leuven, Minderbroedersstraat 12, B-3000, BELGIUM
REFERENCE 2 (bases 1 to 217)
Delabie,J. and Tierens,A.
AUTHORS Evidence for clonal expansion and somatic hypermutations of the marginal zone B cells in the lymph node and spleen
JOURNAL Unpublished
FEATURES
source location/Qualifiers
1..217
/organism="Homo sapiens"
/db_xref="taxon:9606"
/rearranged
1..>217
/codon_start=2
/product="immunoglobulin heavy chain variable region"
/protein_id="CAI2716.1"
/db_xref="GI:2951654"
/translation="FTFSDYVMSWIRQAPGKGLEMLSDSSSGTSIYVDSVKGRFTI SRDNKNSVYLQMNSLRAEDTAVYHCAR"
1..>217
/product="immunoglobulin heavy chain variable region"
BASE COUNT 58 a 55 c 59 g 45 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 217;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 14 GACTACTACATGAGC 28

RESULT 4
AF017465
LOCUS 223 bp mRNA linear PRI 18-OCT-1997
DEFINITION Homo sapiens isolate case 5 rearranged immunoglobulin heavy chain variable region (IGH) mRNA, partial cds.
ACCESSION AF017465
VERSION AF017465.1 GI:2547109
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 223)
Delabie,J. and Pittaluga,S.
AUTHORS Direct Submission
JOURNAL Submitted (08-AUG-1997) Pathology, University of Leuven, Minderbroedersstraat 12, Leuven B-3000, Belgium
FEATURES
source location/Qualifiers
1..223
/organism="Homo sapiens"
/isolate="case 5"
/db_xref="taxon:9606"
1..>223
/gene="IGH"
1..>223
/gene="IGH"
/codon_start=2
/product="rearranged immunoglobulin heavy chain variable region"
/protein_id="AAB81258.1"

/db_xref="GI:2547110"
/translation="SGFTFSDYYMSWIRQAPGKGLEWVSYISSSSYTNYADSVKGRF
TISRDNAKNSLYIQMNSLRADTAVYYCAR"
BASE COUNT 61 a 54 c 59 g 49 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 20 GACTACTACATGAGC 34

RESULT 5
HSJ227734 223 bp DNA linear PRI 10-MAR-1998
LOCUS Homo sapiens DNA for rearranged immunoglobulin heavy chain gene,
DEFINITION 7-9VH3.
ACCESSION AJ227734 GI:2951645
VERSION AJ227734.1
KEYWORDS heavy chain; immunoglobulin; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 223)
AUTHORS Delabie,J.
TITLE Direct Submission
JOURNAL Submitted (03-MAR-1998) Delabie J., Pathology, University of
Leuven, Minderbroedersstraat 12, B-3000, BELGIUM
2 (bases 1 to 223)
REFERENCE Delabie,J. and Tierens,A.
AUTHORS Evidence for clonal expansion and somatic hypermutations of the
TITLE marginal zone B cells in the lymph node and spleen
JOURNAL Unpublished
FEATURES location/Qualifiers
source 1..223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/rearranged
<1..>223
/codon_start=2
/product="immunoglobulin heavy chain variable region"
/protein_id="CAI12712.1"
/db_xref="GI:2951646"
/translation="SGFTFSDYYMSWIRQAPGKGLEWVSYISSSSYTNYADSVKGRF
TISRDNAKNSLYIQMNSLRADTAVYYCAR"
<1..>223
V_region /product="immunoglobulin heavy chain variable region"
BASE COUNT 61 a 54 c 59 g 49 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 223;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 20 GACTACTACATGAGC 34

RESULT 6
HSA306240 246 bp DNA linear PRI 10-JUL-2001
LOCUS Homo sapiens partial 2c-18m3 gene for immunoglobulin heavy chain
DEFINITION variable region.
ACCESSION AJ306240
VERSION AJ306240.1 GI:12641742
KEYWORDS 2c-18m3 gene; immunoglobulin heavy chain; variable region.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
JOURNAL
Thesis (2000) Department of Medizinische Klinik,
Universitaetskliniken des Saarlandes, Homburg, Germany
2 (bases 1 to 246)
Voswinkel,J., Weisgerber,K., Pfreundschuh,M. and Gause,A.
B lymphocyte involvement in ankylosing spondylitis: the heavy chain
variable segment gene repertoire of B lymphocytes from germinal
center-like foci in the synovial membrane indicates antigen
selection
JOURNAL Arthritis Res. 3 (3), 189-195 (2001)
MEDLINE 21196429
PUBMED 11299060
REFERENCE 3 (bases 1 to 246)
AUTHORS Weisgerber,K.
JOURNAL Direct Submission
Submitted (06-JAN-2001) Weisgerber K., Universitaetskliniken des
Saarlandes, Onkologisches Labor, 66424 Homburg, GERMANY
FEATURES location/Qualifiers
source 1..246
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="Synovial membrane"
/rearranged
1..246
/gene="2C-18m3"
<1..>246
/product="immunoglobulin heavy chain variable region"
V_region /gene="2C-18m3"
BASE COUNT 65 a 59 c 68 g 54 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 246;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 19 GACTACTACATGAGC 33

RESULT 7
AF297164 255 bp DNA linear PRI 04-DEC-2000
LOCUS Homo sapiens clone 5-4.3.10A immunoglobulin heavy chain variable
DEFINITION region gene, partial cds.
ACCESSION AF297164
VERSION AF297164.1 GI:11528307
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 255)
AUTHORS Lee,J. and Lipsky,P.E.
TITLE The V lambda J lambda repertoire in human fetal spleen: evidence
for positive selection and extensive receptor editing
JOURNAL Unpublished
2 (bases 1 to 255)
REFERENCE Lee,J. and Lipsky,P.E.
AUTHORS Direct Submission
JOURNAL Submitted (18-AUG-2000) Internal Medicine and Harold C. Simmons
Arthritis Research Center, University of Texas Southwestern Medical
Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA
FEATURES location/Qualifiers
source 1..255
/organism="Homo sapiens"
/isolate="FG5"
/db_xref="taxon:9606"
/chromosome="14"
/clone="5-4.3.10A"

/cell_type="IgM+ B lymphocyte"
 /tissue_type="spleen"
 /dev_stage="fetus; 18 wk gestation"
 /rearranged
 <1. .>255
 /product="immunoglobulin heavy chain variable region"
 <1. .>255
 /note="3-11"
 /codon_start=1
 /product="immunoglobulin heavy chain variable region"
 /protein_id="AAG37215.1"
 /db_xref="GI:11528308"
 /translation="ASGFTFSDDYMSWIRQAPGKGLWVSYSISSGSTIYADSVKGR
 FTISRDNKNSLYLQMSNLRADTAVYYCARAQLGDADFIDW"

mRNA
 CDS
 BASE COUNT
 ORIGIN
 63 a
 61 c
 72 g
 59 t

Query Match
 Best Local Similarity
 100.0%;
 Score 15;
 DB 9;
 Length 255;
 Pred. No. 9.8e+02;
 Matches
 15;
 Conservative
 0;
 Mismatches
 0;
 Indels
 0;
 Gaps
 0;

QY
 1 GACTACTACATGAGC 15
 |||||
 22 GACTACTACATGAGC 36
 Db

RESULT 8
 HSEM112V3
 LOCUS
 HSEM112V3
 DEFINITION
 H.sapiens rearranged Ig heavy chain variable region (EM2112VH3).
 Y12032
 Y12032.1
 GI:1914732
 D-region; immunoglobulin heavy chain; J-region;
 V-region.
 Homo sapiens.
 SOURCE
 Homo sapiens
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 258)
 Foerste,N., Gellrich,S., Golembowski,S., Rutz,S., Audring,H.,
 Sterry,W. and Jahn,S.
 Analysis of VH genes rearranged in single B cells contained in
 dermal infiltrates of patients with Mycosis fungoides
 Unpublished
 2 (bases 1 to 258)
 Jahn,S.
 JOURNAL
 Direct Submission
 TITL
 Submitted (21-MAR-1997) S. Jahn, University Hospital Charite,
 Department of Dermatology, Schumannstrasse 20/21, 10117 Berlin, FRG
 JOURNAL
 Location/Qualifiers
 FEATURES
 source
 1.258
 /organism="Homo sapiens"
 /isolate="patient EM"
 /db_xref="taxon:9606"
 /chromosome="14"
 /cell_type="B lymphocyte"
 /tissue_type="skin biopsy"
 /dev_stage="adult"
 /rearranged
 /note="single B cell in a mycosis fungoides infiltrate"
 1.258
 /gene="EM2112VH3"
 <1. .>258
 /gene="EM2112VH3"
 /codon_start=1
 /product="immunoglobulin heavy chain"
 /protein_id="CAA72763.1"
 /db_xref="GI:1914733"
 /translation="SGFTFSDDYMSWIRQAPGKGLWVSYSISDNFVSADSVKGRF
 TISRDNKNSVFLHMSNLRADTAVYYCARGRAPRPEDYW"

BASE COUNT
 ORIGIN
 66 a
 66 c
 67 g
 59 t

Query Match
 Best Local Similarity
 100.0%;
 Score 15;
 DB 9;
 Length 258;
 Pred. No. 9.8e+02;
 Matches
 15;
 Conservative
 0;
 Mismatches
 0;
 Indels
 0;
 Gaps
 0;

QY
 1 GACTACTACATGAGC 15
 |||||
 Db
 19 GACTACTACATGAGC 33

RESULT 9
 HSA406644
 LOCUS
 HSA406644
 DEFINITION
 Homo sapiens partial VH3-11 gene for rearranged immunoglobulin
 heavy chain variable region, isolate case4-PTGC1-171.
 AJ406644
 AJ406644.1
 GI:10636712
 immunoglobulin heavy chain; variable region; VH3-11 gene.
 human.
 SOURCE
 Homo sapiens
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 261)
 Brauninger,A., Yang,W., Wacker,H.H., Rajewsky,K., Kuppers,R. and
 Hansmann,M.L.
 B-cell development in progressively transformed germinal centers:
 similarities and differences compared with classical germinal
 centers and lymphocyte-predominant Hodgkin disease
 Blood 97 (3), 714-719 (2001)
 21093792
 MEDLINE
 11157489
 PUBMED
 2 (bases 1 to 261)
 Brauninger,A.
 JOURNAL
 Direct Submission
 TITL
 Submitted (21-SEP-2000) Brauninger A., Pathology, University of
 Frankfurt, Theodor Stern Kai 7, 60590, GERMANY
 JOURNAL
 Location/Qualifiers
 FEATURES
 source
 1.261
 /organism="Homo sapiens"
 /isolate="case4-PTGC1-171"
 /db_xref="taxon:9606"
 /rearranged
 1.261
 /gene="VH3-11"
 <1. .>261
 /gene="VH3-11"
 /codon_start=1
 /product="immunoglobulin heavy chain variable region"
 /protein_id="CAC10708.1"
 /db_xref="GI:10636713"
 /translation="SGFTFSDDYMSWIRQAPGKGLWVSYSISSGSTIYADSVKGRF
 TISRDNKNSLYLQMSNLRADTAVYYCARDVRPLWLHYFDYW"

QY
 1 GACTACTACATGAGC 15
 |||||
 Db
 19 GACTACTACATGAGC 33

BASE COUNT
 ORIGIN
 66 a
 62 c
 71 g
 62 t

Query Match
 Best Local Similarity
 100.0%;
 Score 15;
 DB 9;
 Length 261;
 Pred. No. 9.8e+02;
 Matches
 15;
 Conservative
 0;
 Mismatches
 0;
 Indels
 0;
 Gaps
 0;

QY
 1 GACTACTACATGAGC 15
 |||||
 Db
 19 GACTACTACATGAGC 33

RESULT 10
 HSA415895
 LOCUS
 HSA415895
 DEFINITION
 Homo sapiens partial IGVH3-8 gene for immunoglobulin heavy chain
 variable region, donor MF, cell14.
 AJ415895
 AJ415895.1
 GI:16117163

KEYWORDS	IGHV gene; immunoglobulin heavy chain; variable region.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1 (bases 1 to 273)				
TITLE	Brauninger,A., Goossens,T., Rajewsky,K. and Kuppers,R. Regulation of immunoglobulin light chain gene rearrangements during early B cell development in the human				
JOURNAL	Eur. J. Immunol.				
REFERENCE	2 (bases 1 to 273)				
AUTHORS	Goossens,T., Brauninger,A., Klein,U., Kuppers,R. and Rajewsky,K. Receptor revision plays no major role in shaping the receptor repertoire of human memory B cells after the onset of somatic hypermutation				
JOURNAL	Eur. J. Immunol.				
REFERENCE	3 (bases 1 to 273)				
AUTHORS	Brauninger,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (21-SEP-2001) Brauninger A., Pathology, University of Frankfurt, Theodor Stern Kai 7, 60590, GERMANY				
FEATURES	Location/Qualifiers				
source	1. .273				
	/organism="Homo sapiens"				
	/isolate="donor MF"				
	/db_xref="taxon:9606"				
	/clone="cell14"				
	/cell_type="B cell"				
	/rearranged				
	/note="naive lambda expressing"				
gene	1. .273				
	/gene="IGHV3-8"				
V_region	<1. .>273				
	/gene="IGHV3-8"				
BASE COUNT	/product="immunoglobulin heavy chain variable region"				
ORIGIN	72 a 61 c 78 g 62 t				
Query Match	100.0%; Score 15; DB 9; Length 273;				
Best Local Similarity	100.0%; Pred. No. 9.8e+02;				
Matches	15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 GACTACTACATGAGC 15				
Db	1 GACTACTACATGAGC 15				
RESULT 11					
HS277321	278 bp DNA linear PRI 08-NOV-1996				
LOCUS					
DEFINITION	H.sapiens gene for immunoglobulin heavy chain variable region, VH				
ACCESSION	Z77321				
VERSION	Z77321.1 GI:1465675				
KEYWORDS	immunoglobulin heavy chain; immunoglobulin superfamily; variable region.				
SOURCE	Homo sapiens.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1 (bases 1 to 278)				
AUTHORS	Kanzler,H., Kuipers,R., Hansmann,M.L. and Rajewsky,K. Hodgkin and Reed/Sternberg cells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled)				
TITLE	germinal center B cells				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 278)				
AUTHORS	Kanzler,H.				
TITLE	Direct Submission				
JOURNAL	Submitted (18-JUL-1996) Kanzler H., Institute for Genetics at the University of Cologne, Weyertal 121, Cologne 50931, FRG				
REFERENCE	3 (bases 1 to 278)				
AUTHORS	Kanzler,H., Kuppers,R., Hansmann,M.L. and Rajewsky,K.				

TITLE	Hodgkin and Reed-Sternbergcells in Hodgkin's disease represent the outgrowth of a dominant tumor clone derived from (crippled) germinal center B cells			
JOURNAL	J. Exp. Med. 184 (4), 1495-1505 (1996)			
MEDLINE	97033523			
PUBMED	88792220			
FEATURES	Location/Qualifiers			
source	1..278			
	/organism="Homo sapiens"			
	/isolate="patient 8"			
	/db_xref="taxon:9606"			
	/cell_type="Hodgkin-RS"			
	/tissue_type="Hodgkin's lymphoma"			
	/rearranged			
gene	1..278			
	/gene="VH (V3-11)"			
V_region	1..278			
	/gene="VH (V3-11)"			
	/note="immunoglobulin heavy chain variable region"			
BASE COUNT	69 a	64 c	82 g	63 t
ORIGIN				
Query Match	100.0%; Score 15; DB 9; length 278;			
Best local Similarity	100.0%; Pred. No. 9.7e+02;			
Matches	15; Conservative	0; Mismatches	0; Indels	0; Gaps
Qy	1 GACTACTACATGAGC	15		
Db	19 GACTACTACATGAGC	33		
RESULT 12				
AF077461	279 bp DNA linear PRI 04-NOV-1998			
LOCUS	Homo sapiens clone BF3N2g3E09 immunoglobulin heavy chain variable region gene, partial cds.			
DEFINITION	AF077461			
ACCESSION	AF077461.1 GI:3832595			
VERSION				
KEYWORDS	Homo sapiens.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	Brezinschek,H.P., Dorner,T., Farner,N.L., Brezinschek,R.I. and Lipsky,P.E.			
AUTHORS	The Influence of CD40-CD154 Interactions on the Expressed Human Variable Heavy Chain Repertoire: Analysis of VH Genes Expressed by Individual B Cells of a Patient with X-Linked Hyper-Igm Syndrome			
TITLE	Unpublished			
JOURNAL	2 (bases 1 to 279)			
REFERENCE	Brezinschek,H.P., Dorner,T., Farner,N.L., Brezinschek,R.I. and Lipsky,P.E.			
AUTHORS	Direct Submission			
TITLE	Submitted (23-OCT-1998) Internal Medicine, UT Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA			
JOURNAL	Location/Qualifiers			
FEATURES	1..279			
source	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/chromosome="14"			
	/clone="BF3N2g3E09"			
	/cell_type="individual B cell1"			
	/rearranged			
	/note="from patient with X-linked hyper-Igm syndrome"			
	<1..>279			
	/note="VH(D)JH"			
	/codon_start=1			
	/product="immunoglobulin heavy chain variable region"			
	/protein_id="AAC70819.1"			
	/db_xref="GI:3832596"			
	/translation="VKPGSLRLSCAASGFTESDYMSMIRQAPGKLEWYSISSGSGSTIYVADSVKGRFTISRDNAKNSLYLQMSLRAEDTAVYICARDIGVWG"			
CDS				

BASE COUNT 66 a 71 c 81 g 61 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15
|||||
Db 58 GACTACTACATGAGC 72
RESULT 13
AY013309 279 bp DNA linear PRI 03-DEC-2001
LOCUS Homo sapiens clone FM1-3.3.12A immunoglobulin heavy chain variable
DEFINITION region gene, partial cds.
ACCESSION AY013309
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE
AUTHORS Lee, J. and Lipsky, P.E.
TITLE VH gene replacement in human fetal B cells: evidence for immunoglobulin heavy chain receptor editing
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 279)
AUTHORS Lee, J. and Lipsky, P.E.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-2000) Internal Medicine and Harold C. Simmons
Arthritis Research Center, University of Texas Southwestern Medical
Center, 5323 Harry Hines Blvd., Dallas, TX 75235, USA
FEATURES
Source
1. 279
/organism="Homo sapiens"
/isolate="FM1"
/db_xref="taxon:9606"
/chromosome="14"
/clone="FM1-3.3.12A"
/cell_type="IgM+ B lymphocyte"
/tissue_type="bone marrow"
/dev_stage="fetal; 18 weeks gestation"
/rearranged
/note="donor VH3-8; recipient 3-11"
<1..>279
/product="immunoglobulin heavy chain variable region"
<1..>279
/note="CDR2 replacement"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="AAG45480.1"
/db_xref="GI:17220531"
/translation="LRLSCASGFTFSDDYMSWIRQAPGKLEWVSYISSSSSYTNY
ADSVKGRFTISRDNAKNSLYLQMNSLRADDTAVVYCARHDHYGSGSFDYW"
BASE COUNT 70 a 67 c 77 g 65 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 279;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15
|||||
Db 40 GACTACTACATGAGC 54
RESULT 14
AY003824 283 bp mRNA linear PRI 25-SEP-2001
LOCUS Homo sapiens isolate sy-3M/9-H1 immunoglobulin mu heavy chain
DEFINITION

variable region mRNA, partial cds.
ACCESSION AY003824
VERSION AY003824.1 GI:15777397
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 283)
Yavuz, S., Grammer, A.C., Yavuz, A.S., Nanki, T. and Lipsky, P.E.
Comparative characteristics of mu chain and alpha chain transcripts
expressed by individual tonsil plasma cells
Mol. Immunol. 38 (1), 19-34 (2001)
JOURNAL 21376478
MEDLINE 11483207
PUBMED
REFERENCE 2 (bases 1 to 283)
AUTHORS Yavuz, S., Nanki, T., Yavuz, A.S. and Lipsky, P.E.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-2000) Autoimmunity Branch, National Institute of
Arthritis and Musculoskeletal and Skin Diseases, 9000 Rockville Pike
Bldg. 9 1N124, Bethesda, MD 20892, USA
FEATURES
Source
1. 283
/organism="Homo sapiens"
/isolate="sy-3M/9-H1"
/db_xref="taxon:9606"
/cell_type="plasma cell"
<1..>283
/note="VH3 family, 3-11/DP-35"
/codon_start=1
/product="immunoglobulin mu heavy chain variable region"
/protein_id="AAK11793.1"
/db_xref="GI:15777398"
/translation="SDYMSWIRQAPGKLEWVSYISSGSTIYVANSVKGRFTISR
NAKNSLYLQMNSLRADDTAVVYCATDYSGRVRLFLYWGQTLVTVSSG"
BASE COUNT 74 a 72 c 76 g 61 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 283;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15
|||||
Db 4 GACTACTACATGAGC 18
RESULT 15
HUMIGL2E5 291 bp mRNA linear PRI 06-JAN-1995
LOCUS Human (cell line: L2E5) Ig rearranged H-chain mRNA V-region,
DEFINITION partial cds.
ACCESSION L043325
VERSION L043325.1 GI:186055
KEYWORDS V-region; immunoglobulin heavy chain V-region.
SOURCE Homo sapiens fetus liver cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE
AUTHORS Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 291)
Hillson, J.L., Opplinger, I.R., Sasso, E.H., Milner, E. and Wener, M.H.
The emerging B cell repertoire: Influence of developmental stage
and interindividual variation
J. Immunology 149, 3741-3752 (1992)
JOURNAL 2 (sites)
Hillson, J.L., Karr, N.S., Opplinger, I.R., Mannik, M. and Sasso, E.H.
The structural basis of germline-encoded VH3 immunoglobulin binding
to staphylococcal protein A
J. Exp. Med. 178 (1), 331-336 (1993)
JOURNAL 93301610
MEDLINE 8315388
PUBMED
FEATURES
Source
1. 291
Location/Qualifiers

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-13

Perfect score: 15

Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT: *
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT: *
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT: *
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT: *
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT: *
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT: *
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT: *
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT: *
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT: *
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT: *
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT: *
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT: *
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT: *
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT: *
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT: *
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT: *
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT: *
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT: *
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT: *
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT: *
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT: *
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT: *
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT: *
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	20	AA332393
2	15	100.0	294	21	AA246860
3	15	100.0	329	24	ABA94229
4	15	100.0	348	22	AAS03478
5	15	100.0	351	24	AAL43586
6	15	100.0	354	22	AAS03470
7	15	100.0	366	20	AA231651
8	15	100.0	369	22	AAS03533
9	15	100.0	715	16	AA078949

10	14.2	94.7	379	21	AAC00031	Human secreted pro
11	14	93.3	363	22	AAS03474	DNA encoding anti-
12	14	93.3	606	16	AA078960	Human immunoglobul
13	14	93.3	1422	24	ABK63583	Rat sequence diffe
14	14	93.3	2751	24	AAD31625	Arabidopsis thalia
15	14	93.3	3236	24	AAD31644	Arabidopsis thalia
16	14	93.3	8002	24	AAD31629	Arabidopsis thalia
17	13.4	89.3	21	22	AAS07200	RT-PCR down-primer
18	13.4	89.3	228	24	ABN26925	Human ORFX polynuc
19	13.4	89.3	348	22	AAS03446	DNA encoding anti-
20	13.4	89.3	349	22	AAS03504	DNA encoding anti-
21	13.4	89.3	365	22	AAS03447	Human immune/haema
22	13.4	89.3	366	22	AAS03447	DNA encoding anti-
23	13.4	89.3	366	22	AAS03526	DNA encoding anti-
24	13.4	89.3	368	22	AAK83677	Human immune/haema
25	13.4	89.3	372	18	AAT58268	Lead binding Mab 2
26	13.4	89.3	372	22	AAS03448	DNA encoding anti-
27	13.4	89.3	467	22	ABA51889	Human foetal liver
28	13.4	89.3	467	22	ABA21709	Probe #175 for gen
29	13.4	89.3	467	22	AAK00186	Human brain expres
30	13.4	89.3	467	22	AAK25626	Human bone marrow
31	13.4	89.3	467	22	AAI10253	Probe #186 for gen
32	13.4	89.3	467	22	AAI31500	Probe #181 used to
33	13.4	89.3	467	22	AAI00190	Human genome-deriv
34	13.4	89.3	467	24	ABS00195	Human secreted pro
35	13.4	89.3	529	21	AAC08128	Aspergillus oryzae
36	13.4	89.3	674	21	AAF13480	Aspergillus oryzae
37	13.4	89.3	675	21	AAF12980	Trichoderma reesei
38	13.4	89.3	729	21	AAF15002	P4-2 single chain
39	13.4	89.3	777	24	AAS97135	Artificial synthe
40	13.4	89.3	1080	20	AAZ21160	Artificial synthe
41	13.4	89.3	1086	20	AAZ21157	Artificial synthe
42	13.4	89.3	1113	20	AAZ21159	Artificial synthe
43	13.4	89.3	1167	24	ABN59693	Novel human coding
44	13.4	89.3	1446	20	AAZ23804	Murine Kv6.2 CDNA.
45	13.4	89.3	1571	21	AAC77699	Human cancer assoc

ALIGNMENTS

RESULT 1	AA332393	standard; DNA; 15 BP.
ID	AA332393	
XX	AA332393;	
AC	17-JUN-1999	(first entry)
DT		
XX		
DE	Ab2 variable heavy (VH) chain CDR1 encoding DNA.	
XX		
KW	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;	
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;	
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;	
KW	myelodysplasia; myelotoxic chemotherapy; leukaemia; tumour; MUSK; CDR;	
KW	neuromuscular; muscular dystrophy; complementarity determining region;	
XX	variable heavy chain; variable light chain; VH; VL; SS.	
OS	Homo sapiens.	
XX		
PN	WO9910494-A2.	
PN		
PD	04-MAR-1999.	
XX		
PF	21-AUG-1998;	98WO-US17364.
XX		
PR	25-AUG-1997;	97US-0918148.
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;	
XX		
DR	WPI; 1999-204666/17.	

PI Rodriguez M, Miller DJ, Pease LR;
 XX
 DR WPI; 2002-066596/09.
 DR P-PSDB; ABB07182.
 XX
 PT Novel neuromodulatory agent (a human IGM monoclonal antibody),
 PT promoting neurite outgrowth, regeneration, remyelination and
 PT neuroprotection in central nervous system, useful to treat
 PT post-infectious encephalomyelitis -
 XX
 PS Claim 43; Fig 41; 219pp; English.
 XX
 CC The invention provides a neuromodulatory agent (I) capable of promoting
 CC neurite outgrowth, regeneration, remyelination and neuroprotection in
 CC central nervous system (CNS). (I) is capable of inducing remyelination,
 CC promoting cellular proliferation of glial cells, and promoting Ca2+
 CC signaling with oligodendrocytes. An humanised antibody to (I) can be
 CC selected from antibody SHIGM22 (LYM 22), ebvHlgM Msi19D10, ebv HlgM
 CC CB2bG8, AKJR4, CB2IE12, CB2IE7 or Msi19E5. (I) is useful for stimulating
 CC remyelination of CNS axons, stimulating proliferation of glial cells in
 CC CNS axons, or treating demyelinating disease of CNS in a mammal in need
 CC of such therapy. (I) is capable of binding to structures and cells within
 CC CNS. (I) is preferably useful for treating a demyelinating disease of CNS
 CC of a mouse infected with Strain DA of Theiler's murine encephalomyelitis
 CC (TMEV) or for treating a human being having multiple sclerosis, or a post-
 CC human or domestic animal with a viral demyelinating disease, or a post-
 CC neural disease of CNS. (I) is also useful for an in vitro method of
 CC stimulating the proliferation of glial cells from mixed cell culture.
 CC (I) is also useful for stimulating remyelination of CNS axons. The
 CC antibodies are useful for preventing infection by a bacterium, virus or
 CC like pathogen that causes demyelination or other neurodegenerative
 CC condition in a subject. Methods where (I) is administered to a patient
 CC are useful for treating multiple sclerosis, Parkinson's disease,
 CC Alzheimer's disease, amyotrophic lateral sclerosis (ALS), a viral
 CC demyelinating disease, CNS diseases, and other conditions in the CNS
 CC where nerves are damaged as by trauma. The present sequence represents
 CC the CB2IE7 heavy chain variable region nucleotide sequence.
 XX
 SQ Sequence 329 BP; 81 A; 81 C; 96 G; 71 T; 0 other;
 Query Match 100.0%; Score 15; DB 24; Length 329;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACTACTACATGAGC 15
 XX ||||||||||||||||
 DB 66 GACTACTACATGAGC 80
 RESULT 4
 AAS03478
 ID AAS03478 standard; cDNA; 348 BP.
 AC AAS03478;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 65.
 XX
 KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 KW heart disease; complementarity determining region; CDR; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200127279-A1.
 XX
 PD 19-APR-2001.
 XX
 PF 11-OCT-2000; 2000WO-GB03900.
 XX
 PR 12-OCT-1999; 99US-0158812.
 XX
 PA (CAMB-) CAMBRIDGE ANTIBODY TECHNOLOGY.

XX
 PI Edwards BM, Main SH, Vaughan TJ;
 XX
 DR WPI; 2001-282031/29.
 DR P-PSDB; AAU02578.
 XX
 PT Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -
 XX
 PS Disclosure; Page 141; 182pp; English.
 XX
 CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radioisotope, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.
 XX
 SQ Sequence 348 BP; 79 A; 91 C; 108 G; 70 T; 0 other;
 Query Match 100.0%; Score 15; DB 22; Length 348;
 Best Local Similarity 100.0%; Pred. No. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GACTACTACATGAGC 15
 XX ||||||||||||||||
 DB 91 GACTACTACATGAGC 105
 RESULT 5
 AAL43586
 ID AAL43586 standard; DNA; 351 BP.
 AC AAL43586;
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Dig3 antibody heavy chain coding sequence.
 XX
 KW 26-10 scFv antibody light chain; complementarity determining region 3;
 KW CDR3; binding protein production; catalytic protein production; ds;
 KW ligand structure determination; antibody isolation; dig3 heavy chain.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..351
 FT /*tag= a
 FT /partial
 FT /product= "Dig3 heavy chain"
 FT /note= "No start or stop codon is given"
 XX
 PN WO200234886-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 26-OCT-2001; 2001WO-US46795.
 XX
 PR 27-OCT-2000; 2000US-0699023.
 XX
 PA (TEXA) UNIV TEXAS SYSTEM.
 XX

PI Chen G, Hayhurst A, Thomas JG, Iverson BL, Georgiou G;
XX
DR WPI; 2002-500078/53.
DR P-PSDB; AAO15187.
XX
PT Obtaining bacterium having nucleic acid encoding binding protein that
PT binds target ligand, or a nucleic acid encoding catalytic protein that
PT catalyzes reaction involving target substrate by display-less library
PT screening -
XX
PS Example 4; Page 97; 98pp; English.
XX
CC The invention comprises a method of obtaining a bacterium containing a
CC nucleic acid encoding a binding protein capable of binding a target
CC ligand, or a nucleic acid encoding a catalytic protein that catalyses a
CC reaction involving a target substrate. The method of the invention allows
CC the isolation of a cell expressing a nucleic acid encoding a binding
CC protein capable of binding a target ligand, and consequently the
CC isolation of the nucleic acid. The binding protein produced by the method
CC of the invention is useful in determining the structure of a target
CC ligand. The method of the invention is useful for the de novo isolation
CC of antibodies from large repertoire libraries and for monitoring
CC production during protein manufacturing. The method can also be used for
CC monitoring production of a particular byproduct of a biological reaction.
CC The present DNA sequence encodes the dig3 scFv antibody heavy chain which
CC is used in an example of the invention.
XX
SQ Sequence 351 BP; 77 A; 85 C; 109 G; 80 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 351;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
AC ||||||||||||||
DB 91 GACTACTACATGAGC 105

RESULT 6
AAS03470
ID AAS03470 standard; cDNA; 354 BP.
XX
AC AAS03470;
XX
DT 29-AUG-2001 (first entry)
XX
DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 58.
XX
KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
KW heart disease; complementarity determining region; CDR; ss.
XX
OS Homo sapiens.
XX
PN WO200127279-A1.
XX
PD 19-APR-2001.
XX
PF 11-OCT-2000; 2000WO-GB03900.
XX
PR 12-OCT-1999; 99US-0158812.
XX
PA (CAMB-) CAMBRIDGE ANTIBODY TECHNOLOGY.
XX
PI Edwards BM, Main SH, Vaughan TJ;
XX
DR WPI; 2001-282031/29.
DR P-PSDB; AAU02570.
XX
PT Panel of specific binding members of antibody molecules which bind to
PT whole adipocytes is used in the treatment of obesity and obesity
PT related diseases -
XX
PS Disclosure; Page 136; 182pp; English.

XX
CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
CC and light chain coding sequences of the invention. The antibodies can be
CC used in the treatment of obesity and obesity related diseases. The
CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
CC mass of an obese patient or the antibody can be used as a therapeutic
CC itself. Antibodies binding specifically to adipocytes can be used to
CC activate the immune system to destroy the cells by complement mediated
CC lysis. The antibodies may be labeled with a detectable label such as
CC radiolabel, fluorescent or chemical group and used in methods of
CC diagnosis in human subjects e.g. to determine the presence of adipocyte
CC antigen on the surface of an adipocyte to detect or determine the
CC presence or level of adipocytes in a cell or tissue sample. The
CC antibodies can be used as an alternative means of treatment for obese
CC patients other than undergoing surgery to remove excess fat. Antibodies
CC for different types of fat deposits can also be produced e.g. intra-
CC abdominal fat associated with heart disease.
XX
SQ Sequence 354 BP; 83 A; 84 C; 108 G; 79 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 354;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
AC ||||||||||||||
DB 91 GACTACTACATGAGC 105

RESULT 7
AAZ31651
ID AAZ31651 standard; DNA; 366 BP.
XX
AC AAZ31651;
XX
DT 13-JAN-2000 (first entry)
XX
DE Coding sequence for VH domain CDR of anti-estradiol antibody.
XX
KW Estradiol; complementarity determining region; CDR; estriol-3-sulphate;
KW antibody antigen binding domain; steroid hormone; estriol; testosterone;
KW dihydrotestosterone; progesterone; estriol 3-beta-di-glucuronide;
KW menstrual cycle; hormone replacement therapy; oestrogen secreting tumour;
KW diagnosis; VH domain; ds.
XX
OS Homo sapiens.
XX
PN US5977319-A.
XX
PD 02-NOV-1999.
XX
PF 21-OCT-1997; 97US-0958201.
XX
PR 21-OCT-1996; 96US-0028897.
XX
PA (CAMB-) CAMBRIDGE ANTIBODY TECHNOLOGY.
XX
PI Johnson KS, Pope AR, Pritchard K, Williams AJ;
XX
DR WPI; 1999-619713/53.
DR P-PSDB; AAY43253.
XX
PT New specific binding partners for estradiol, used for monitoring
PT estradiol levels during the menstrual cycle, in hormone replacement
PT therapy and for diagnosing oestrogen secreting tumours -
XX
PS Disclosure; Column 21-22; 26pp; English.
XX
CC This sequence encodes a VH domain complementarity determining region
CC (CDR) from an antibody specific for estradiol. The invention relates to
CC specific binding members (sbp) comprising a polypeptide that comprises an
CC antibody antigen binding domain (AABD) which has a dissociation constant
CC of less than 1.0x10^-8M for estradiol, and a dissociation constant of at

CC least 500-fold higher for the steroid hormones selected from estradiol, testosterone, dihydrotestosterone, progesterone, estradiol-3-sulphate and estradiol 3-beta-di-glucuronide, where the polypeptide comprises an antibody VH domain. The sbps can be used in an immunoassay for determining the presence or absence of estradiol in a sample. They can be used for monitoring estradiol levels, e.g. during the menstrual cycle, in hormone replacement therapy and for diagnosing oestrogen secreting tumours. The sbps can provide for discrimination between estradiol and other related steroids.

CC CC

XX SQ Sequence 366 BP; 80 A; 98 C; 110 G; 77 T; 1 other;

Query Match 100.0%; Score 15; DB 20; Length 366;
 Best Local Similarity 100.0%; Pred. NO. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
 |||||
 Db 91 GACTACTACATGAGC 105

RESULT 8
 AAS03533
 ID AAS03533 standard; cDNA; 369 BP.
 AC AAS03533;
 XX
 DT 29-AUG-2001 (first entry)
 XX
 DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 116.
 XX
 KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 KM heart disease; complementarity determining region; CDR; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200127279-A1.
 XX
 PD 19-APR-2001.
 XX
 PF 11-OCT-2000; 2000WO-GB03900.
 XX
 PR 12-OCT-1999; 99US-0158812.
 XX
 PA (CAMB-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 XX
 PI Edwards BM, Main SH, Vaughan TJ;
 XX
 DR WPI; 2001-282031/29.
 DR P-PSDB; AAU02633.
 XX
 PT Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -
 XX
 PS Disclosure; Page 175; 182pp; English.
 XX
 CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.

XX SQ Sequence 369 BP; 81 A; 90 C; 119 G; 79 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 369;
 Best Local Similarity 100.0%; Pred. NO. 66;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
 |||||
 Db 91 GACTACTACATGAGC 105

RESULT 9
 AAQ78949
 ID AAQ78949 standard; DNA; 715 BP.
 AC AAQ78949;
 XX
 DT 01-AUG-1995 (first entry)
 XX
 DE Human immunoglobulin Vh gene #11.
 XX
 KW Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;
 KM cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 206..653
 FT /tag= a
 FT /product= human immunoglobulin variable heavy chain
 FT intron 252..348
 FT /tag= b
 FT misc_signal 654..656
 FT /tag= c
 FT /note= "miscellaneous signal, does not conform to
 terminator or splice site sequence"

XX PN WO9426895-A.
 XX
 PD 24-NOV-1994.
 XX
 PF 10-MAY-1993; 93WO-JP00603.
 XX
 PR 10-MAY-1993; 93WO-JP00603.
 XX
 PA (NISB) JAPAN TOBACCO INC.
 XX
 PI Honjo T, Matsuda F;
 XX
 DR WPI; 1995-006791/01.
 DR P-PSDB; AAR66305.
 XX
 PT DNA fragment comprising human immunoglobulin Vh genes - for the
 PT production of human immunoglobulin in mammalian hosts
 PT
 XX
 PS Claim 20; Page 44-45; 130pp; Japanese.
 XX
 CC A series of genes (AAQ78939-79002) encoding human immunoglobulin
 CC variable heavy chains. The genes were isolated and cloned from a series
 CC of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M118 and M131,
 CC by PCR amplification using primers AAQ78917-38. The genes are subdivided
 CC into 5 families of Vh genes. The fragments cover a region of 800 kb.
 CC The DNA fragments were isolated from high molecular weight DNA from
 CC human placenta. The DNA was partially digested with TaqI restriction
 CC enzyme. The fragments were separated by gel electrophoresis and 35-45 kb
 CC cosmid vector pJB81. The ligation products were in vitro packed and
 CC infected into E.coli 490A. The fragments were then subcloned by colony
 CC hybridisation. The Vh genes and the DNA fragments encoding them are
 CC useful in producing human immunoglobulin in mammalian hosts.
 XX
 SQ Sequence 715 BP; 170 A; 162 C; 209 G; 174 T; 0 other;

Query Match 100.0%; Score 15; DB 16; Length 715;
 Best Local Similarity 100.0%; Pred. No. 69;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
 |||||
 Db 450 GACTACTACATGAGC 464

RESULT 10
 AAC00031
 ID AAC00031 standard; cDNA; 379 BP.

AC AAC00031;
 DT 06-OCT-2000 (first entry)
 DE Human secreted protein 5' EST, SEQ ID NO: 29.
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
 KW gene therapy; chromosome mapping; ss.

OS Homo sapiens.
 XX EPI033401-A2.
 XX 06-SEP-2000.
 XX 21-FEB-2000; 2000EP-0200610.
 XX 26-FEB-1999; 99US-0122487.
 PA (GEST) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;
 DR WPI; 2000-500381/45.
 DR P-PSDB; AAG00025.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
 PS Claim 1; SEQ ID 29; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from
 CC mRNAs encoding secreted proteins. An ORF has been identified within the
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
 CC derived from 30 different tissues. EST sequences usually correspond
 CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
 CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
 CC well suited for isolating cDNA sequences derived from the 5' ends of
 CC mRNAs and even in those cases where longer cDNA sequences have been
 CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
 CC mRNAs with intact 5' ends and can therefore be used to obtain full length
 CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
 CC gene therapy and chromosome mapping procedures. They are used to obtain
 CC upstream regulatory sequences and to design expression and secretion
 CC vectors.

SQ Sequence 379 BP; 83 A; 88 C; 102 G; 85 T; 21 other;

Query Match 94.7%; Score 14.2; DB 21; Length 379;
 Best Local Similarity 86.7%; Pred. No. 1.9e+02;
 Matches 13; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 15
 |||||
 Db 227 GACTWCTACATGARC 241

RESULT 11

AAS03474
 ID AAS03474 standard; cDNA; 363 BP.
 AC AAS03474;
 DT 29-AUG-2001 (first entry)
 DE DNA encoding anti-adipocyte monoclonal antibody heavy chain, FAT 62.
 KW Antibody; adipocyte; heavy chain; light chain; obesity; FAT;
 KW heart disease; complementarity determining region; CDR; ss.

OS Homo sapiens.
 XX WO200127279-A1.
 XX 19-APR-2001.
 XX 11-OCT-2000; 2000WO-GB03900.
 XX 12-OCT-1999; 99US-0158812.
 PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
 PI Edwards BM, Main SH, Vaughan TJ;
 DR WPI; 2001-282031/29.
 DR P-PSDB; AAU02574.

PT Panel of specific binding members of antibody molecules which bind to
 PT whole adipocytes is used in the treatment of obesity and obesity
 PT related diseases -
 PS Disclosure; Page 139; 182pp; English.

CC AAS03401-AAS03535 represent anti-adipocyte monoclonal antibody heavy
 CC and light chain coding sequences of the invention. The antibodies can be
 CC used in the treatment of obesity and obesity related diseases. The
 CC antibodies can be used to deliver drugs or pro-drugs directly to the fat
 CC mass of an obese patient or the antibody can be used as a therapeutic
 CC itself. Antibodies binding specifically to adipocytes can be used to
 CC activate the immune system to destroy the cells by complement mediated
 CC lysis. The antibodies may be labeled with a detectable label such as
 CC radiolabel, fluorescent or chemical group and used in methods of
 CC diagnosis in human subjects e.g. to determine the presence of adipocyte
 CC antigen on the surface of an adipocyte to detect or determine the
 CC presence or level of adipocytes in a cell or tissue sample. The
 CC antibodies can be used as an alternative means of treatment for obese
 CC patients other than undergoing surgery to remove excess fat. Antibodies
 CC for different types of fat deposits can also be produced e.g. intra-
 CC abdominal fat associated with heart disease.

SQ Sequence 363 BP; 84 A; 91 C; 103 G; 85 T; 0 other;

Query Match 93.3%; Score 14; DB 22; Length 363;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTACTACATGAGC 14
 |||||
 Db 91 GACTACTACATGAGC 104

RESULT 12
 AAQ078960
 ID AAQ078960 standard; DNA; 606 BP.

AC AAQ078960;
 DT 03-AUG-1995 (first entry)
 DE Human immunoglobulin Yh gene #22.

Accession	Gene	Location/Qualifiers
XX	Primer; PCR; amplify; human; immunoglobulin; variable; heavy chain;	
KM	cosmid; placenta; vector; pJB81; E.coli; mammalian; ds.	
XX		
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	85..544
FT		/*tag= a
FT	intron	/product= human immunoglobulin variable heavy chain
FT		131..233
FT		/*tag= b
FT	misc_signal	428..430
FT		/*tag= c
FT	misc_signal	/note= "non-functional termination codon"
FT		545..547
FT		/*tag= d
FT		/note= "miscellaneous signal, does not conform to termination or splice site sequence"
XX		
PN	WO9426895-A.	
XX		
PD	24-NOV-1994.	
XX		
PF	10-MAY-1993; 93WO-JP00603.	
XX		
PR	10-MAY-1993; 93WO-JP00603.	
XX		
PA	(NISB) JAPAN TOBACCO INC.	
PI	Honjo T, Matsuda F;	
XX		
DR	WPI; 1995-006791/01.	
XX		
DR	P-PSDB; AAR66314.	
XX		
PT	DNA fragment comprising human immunoglobulin Vh genes - for the	
PT	production of human immunoglobulin in mammalian hosts	
XX		
PS	Claim 31; Page 59-60; 130pp; Japanese.	
XX		
CC	A series of genes (AAQ78939-79002) encoding human immunoglobulin	
CC	variable heavy chains. The genes were isolated and cloned from a series	
CC	of cosmid constructs: Y202; Y103; Y21; Y6; Y24; 3-31; M84; M18 and M131,	
CC	by PCR amplification using primers AAQ78917-38. The genes are subdivided	
CC	into 5 families of Vh genes. The fragments cover a region of 800 kb.	
CC	The DNA fragments were isolated from high molecular weight DNA from	
CC	human placenta. The DNA was partially digested with TaqI restriction	
CC	enzyme. The fragments were separated by gel electrophoresis and 35-45 kb	
CC	fractions were collected. The fragments were ligated with ClaI-digested	
CC	cosmid vector pJB81. The ligation products were in vitro packed and	
CC	infected into E.coli 490A. The fragments were then subcloned by colony	
CC	hybridisation. The Vh genes and the DNA fragments encoding them are	
CC	useful in producing human immunoglobulin in mammalian hosts.	
XX		
SO	Sequence 606 BP; 150 A; 128 C; 184 G; 144 T; 0 other;	
QY	Query Match	93.3%; Score 14; DB 16; Length 606;
	Best Local Similarity	100.0%; Pred. NO. 2.5e+02;
	Matches 14; Conservative	0; Mismatches 0; Indels 0; Gaps 0
DB	2 ACTACTACATGAGC 15	
	336 ACTACTACATGAGC 349	
RESULT 13		
ABK63583	standard; cDNA; 1422 BP.	
AC	ABK63583;	
XX		
DT	18-JUN-2002 (first entry)	
XX		
DE	Rat sequence differentially expressed in response to a hepatotoxin #1490.	

XX Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
 KW differential expression; centrilobular necrosis; steatosis.
 XX
 OS Rattus norvegicus.
 XX
 PN W0200210453-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 30-JUL-2001; 2001WO-US23872.
 XX
 PR 31-JUL-2000; 2000US-222040P.
 PR 02-NOV-2000; 2000US-244880P.
 PR 11-MAY-2001; 2001US-290029P.
 PR 15-MAY-2001; 2001US-290645P.
 PR 22-MAY-2001; 2001US-292336P.
 PR 06-JUN-2001; 2001US-295798P.
 PR 13-JUN-2001; 2001US-297457P.
 PR 19-JUN-2001; 2001US-298884P.
 PR 09-JUL-2001; 2001US-303459P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
 XX
 DR WPI; 2002-241625/29.
 XX
 PT Predicting toxic effects of compounds or the progression of these toxic
 PT effects by determining the changes in gene expression in tissues or
 PT cells exposed to the toxin and comparing these to gene expression in
 PT unexposed tissues or cells -
 PS
 PS Claim 1; Seq ID No 1490; 239pp; English.
 XX
 XX The invention relates to methods for predicting toxic effects of
 CC compounds or the progression of these toxic effects by determining the
 CC global changes in gene expression in tissues or cells exposed to the
 CC toxin and comparing these to gene expression in unexposed tissues or
 CC cells. Also included are methods of predicting at least one toxic
 CC effect of a compound or progression of a toxic effect, preferably the
 CC hepatotoxicity of a compound, comprising detecting the level of
 CC expression in a tissue or cell sample exposed to the compound of two or
 CC more genes listed in the specification, where differential expression of
 CC the genes is indicative of at least one toxic effect or progression.
 CC The method can also be used to identify an agent which modulates the
 CC toxic response and predict cellular pathways that a compound modulates
 CC in a cell. The methods utilise a set of at least two probes (on a solid
 CC support in kit form), where each of the probes comprises a sequence that
 CC specifically hybridises to a gene listed in the specification, a computer
 CC system comprising a database containing information identifying the
 CC expression level in a tissue or cell sample exposed to a hepatotoxin of a
 CC set of genes comprising at least two genes listed in the specification,
 CC and a user interface to view the information used to present information
 CC identifying the expression level in a tissue or cell of at least one gene
 CC listed in the specification. The method is useful for elucidating global
 CC changes in gene expression and for identifying toxicity markers in
 CC tissues or cell exposed to a known toxin. The genes may be used as
 CC toxicity markers in drug screening and toxicity assays. The genes and
 CC gene expression information may be used as diagnostic markers for the
 CC prediction or identification of the physiological state of tissue or cell
 CC sample that has been exposed to a compound or agent. Hepatotoxicity
 CC is characterised by centrilobular necrosis and steatosis. The present
 CC sequence is an expressed sequence tag (EST) or cDNA derived from a gene
 CC which is differentially expressed in response to a hepatotoxic agent.
 XX
 XX
 SQ Sequence 1422 BP; 425 A; 305 C; 246 G; 446 T; 0 other;

Query Match 93.3%; Score 14; DB 24; Length 1422;
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 GACTACTACATGAG 14

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	4	US-08-918-148-13 Sequence 13, Appl
2	15	100.0	366	2	US-08-958-201-5 Sequence 5, Appl
3	15	100.0	715	3	US-08-545-809A-11 Sequence 11, Appl
4	14	93.3	606	3	US-08-545-809A-22 Sequence 22, Appl
5	14	93.3	2073	4	US-09-221-017B-404 Sequence 404, App
6	13.4	89.3	372	3	US-08-767-128-37 Sequence 37, Appl
7	13.4	89.3	2004	1	US-08-471-033-18 Sequence 18, Appl
8	13.4	89.3	2004	2	US-08-471-044-18 Sequence 18, Appl
9	13.4	89.3	2004	2	US-08-463-483A-18 Sequence 18, Appl
10	13.4	89.3	2004	2	US-08-471-046A-18 Sequence 18, Appl
11	13.4	89.3	2004	2	US-08-470-566B-18 Sequence 18, Appl
12	13.4	89.3	2004	2	US-08-469-334-18 Sequence 18, Appl
13	13.4	89.3	2004	3	US-09-300-529-18 Sequence 18, Appl
14	13.4	89.3	2576	1	US-08-471-033-35 Sequence 35, Appl
15	13.4	89.3	2576	2	US-08-471-044-35 Sequence 35, Appl
16	13.4	89.3	2576	2	US-08-463-483A-35 Sequence 35, Appl
17	13.4	89.3	2576	2	US-08-471-046A-35 Sequence 35, Appl
18	13.4	89.3	2576	2	US-08-470-566B-35 Sequence 35, Appl
19	13.4	89.3	2576	2	US-08-469-334-35 Sequence 35, Appl
20	13.4	89.3	2576	3	US-09-300-529-35 Sequence 35, Appl
21	13.4	89.3	2655	1	US-08-471-033-17 Sequence 17, Appl
22	13.4	89.3	2655	1	US-08-471-033-26 Sequence 26, Appl
23	13.4	89.3	2655	2	US-08-471-044-17 Sequence 17, Appl
24	13.4	89.3	2655	2	US-08-471-044-26 Sequence 26, Appl
25	13.4	89.3	2655	2	US-08-463-483A-17 Sequence 17, Appl
26	13.4	89.3	2655	2	US-08-463-483A-26 Sequence 26, Appl
27	13.4	89.3	2655	2	US-08-471-046A-17 Sequence 17, Appl

28	13.4	89.3	2655	2	US-08-471-046A-26	Sequence 26, Appl
29	13.4	89.3	2655	2	US-08-470-566B-17	Sequence 17, Appl
30	13.4	89.3	2655	2	US-08-470-566B-26	Sequence 26, Appl
31	13.4	89.3	2655	2	US-08-469-334-17	Sequence 17, Appl
32	13.4	89.3	2655	2	US-08-469-334-26	Sequence 26, Appl
33	13.4	89.3	2655	3	US-09-300-529-17	Sequence 17, Appl
34	13.4	89.3	2655	3	US-09-300-529-26	Sequence 26, Appl
35	13.4	89.3	4031	1	US-08-471-033-49	Sequence 49, Appl
36	13.4	89.3	4031	2	US-08-471-044-49	Sequence 49, Appl
37	13.4	89.3	4031	2	US-08-463-483A-49	Sequence 49, Appl
38	13.4	89.3	4031	2	US-08-471-046A-49	Sequence 49, Appl
39	13.4	89.3	4031	2	US-08-470-566B-49	Sequence 49, Appl
40	13.4	89.3	4031	2	US-08-469-334-49	Sequence 49, Appl
41	13.4	89.3	4031	3	US-09-300-529-49	Sequence 49, Appl
42	13	86.7	464	2	US-08-449-287-9	Sequence 9, Appl
43	13	86.7	464	2	US-08-449-287-11	Sequence 11, Appl
44	13	86.7	777	2	US-08-860-882A-28	Sequence 28, Appl
45	13	86.7	777	4	US-09-011-769A-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-08-918-148-13
; Sequence 13, Application US/08918148A
; Patent No. 6342220
; GENERAL INFORMATION:
; APPLICANT: Adams, Camellia
; APPLICANT: W.
; APPLICANT: Carter, Paul J.
; APPLICANT: Fendly, Brian M.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Agonist Antibodies
; FILE REFERENCE: P0979
; CURRENT APPLICATION NUMBER: US/08/918,148A
; CURRENT FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 79
; SEQ ID NO 13
; LENGTH: 15
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; NAME/KEY: 5E5SCFV VH CDR1
; LOCATION: 1-15
; OTHER INFORMATION:
US-08-918-148-13

Query Match 100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.2;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
Db 1 GACTACTACATGAGC 15

RESULT 2
US-08-958-201-5
; Sequence 5, Application US/08958201
; Patent No. 5977319
; GENERAL INFORMATION:
; APPLICANT: Pope, Anthony R
; APPLICANT: Pritchard, Kevin J
; APPLICANT: Williams, Andrew J
; APPLICANT: Johnson, Kevin S
; TITLE OF INVENTION: Specific binding members for estradiol;
; TITLE OF INVENTION: materials and methods
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall O'Toole Gerstein Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago

STATE: Illinois
COUNTRY: USA
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/958,201
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,897
FILING DATE: 21-OCT-1996
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 2G
FEATURE:
NAME/KEY: CDS
LOCATION: 1..366
US-08-958-201-5

Query Match 100.0%; Score 15; DB 2; Length 366;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
Db 91 GACTACTACATGAGC 105

RESULT 3
US-08-545-809A-11
Sequence 11, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-8906
TELEFAX: 200154
INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:
LENGTH: 715 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-11

Query Match 100.0%; Score 15; DB 3; Length 715;
Best Local Similarity 100.0%; Pred. No. 9.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
Db 450 GACTACTACATGAGC 464

RESULT 4
US-08-545-809A-22
Sequence 22, Application US/08545809A
Patent No. 6096878
GENERAL INFORMATION:
APPLICANT: Honjo, Tasuku
TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
NUMBER OF SEQUENCES: 145
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/545,809A
FILING DATE: 27-MAR-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP93/00603
FILING DATE: 10-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Freeman, John W.
REGISTRATION NUMBER: 29,066
REFERENCE/DOCKET NUMBER: 06501/004001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-8906
TELEFAX: 200154
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 606 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: human lymphoblast
CELL LINE: CGM1
US-08-545-809A-22

Query Match 93.3%; Score 14; DB 3; Length 606;
Best Local Similarity 100.0%; Pred. No. 33;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	2	ACTACTACATGAGC	15
Db	336	ACTACTACATGAGC	349

RESULT 5

```

US-09-221-017B-404
; Sequence 404, Application US/09221017B
; Patent No. 6444799
;
; GENERAL INFORMATION:
;   APPLICANT: Ross, Bruce C.
;   TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;   NUMBER OF SEQUENCES: 1120
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: MORRISON & FOERSTER
;     STREET: 755 PAGE MILL ROAD
;     CITY: Palo Alto
;     STATE: CA
;     COUNTRY: USA
;     ZIP: 94304-1018
;
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: Windows
;   SOFTWARE: FASTSEQ for Windows Version 2.0b
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/221,017B
;   FILING DATE: 23-DEC-1998
;
; CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PP1182
;     FILING DATE: 31-DEC-1997
;
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PP1546
;     FILING DATE: 30-JAN-1998
;
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PP2911
;     FILING DATE: 09-APR-1998
;
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/AU98/01023
;     FILING DATE: 10-DEC-1998
;
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Monroy, Gladys H
;     REGISTRATION NUMBER: 32,430
;     REFERENCE/DOCKET NUMBER: 27340-20021.00
;
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 650-813-5600
;     TELEFAX: 650-494-0792
;
;   TELEX: 706141
;
; INFORMATION FOR SEQ ID NO: 404:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 2073 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: circular
;
;   MOLECULE TYPE: DNA (genomic)
;
;   HYPOTHETICAL: NO
;
;   ANTI-SENSE: UNKNOWN
;
;   ORIGINAL SOURCE:
;     ORGANISM: PORYPHYROMONAS GINGIVALIS
;
;   FEATURE:
;     NAME/KEY: misc_feature
;     LOCATION: 1...2073
;
US-09-221-017B-404

```

Query Match	93.38;	Score 14;	DB 4;	Length 2073;
Best Local Similarity	100.0%;	Pred. No. 38;		
Matches 14; Conservative	0;	Mismatches	0;	Indels 0;
QY	1	GA	CT	ACTACATGAG 14
DB	1828	GA	CT	ACTACATGAG 1841

RESULT 6
US-08-767-128-37
; Sequence 37, Application US/08767128
; Patent No. 6111079

```

1  APPLICANT:  WYLIE, DWANE E.
2  APPLICANT:  LOPEZ, OSVALDO
3  APPLICANT:  MURRAY, PETER JOSEPH
4  APPLICANT:  GOEBEL, PETER
5  TITLE OF INVENTION:  LEAD BINDING POLYPEPTIDES AND
6  TITLE OF INVENTION:  NUCLEOTIDES CODING THEREFORE
7  NUMBER OF SEQUENCES:  46
8  CORRESPONDENCE ADDRESS:
9  ADDRESSEE:  Merchant, Gould, Smith, Edell, Welter & Schmidt
10 STREET:  3100 No. 6111079west Center, 90 South Seventh St
11 CITY:  Minneapolis
12 STATE:  MN
13 COUNTRY:  USA
14 ZIP:  55402
15
16 COMPUTER READABLE FORM:
17 MEDIUM TYPE:  Diskette
18 COMPUTER:  IBM Compatible
19 OPERATING SYSTEM:  DOS
20 SOFTWARE:  FastSeq Version 1.5
21
22 CURRENT APPLICATION DATA:
23 APPLICATION NUMBER:  US/08/767,128
24
25 FILING DATE:
26 CLASSIFICATION:  424
27
28 PRIOR APPLICATION DATA:
29 APPLICATION NUMBER:
30 FILING DATE:  04-DEC-1996
31 CLASSIFICATION:  424
32
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER:  PCT/US96/09258
35 FILING DATE:  05-JUN-1996
36
37 PRIOR APPLICATION DATA:
38 APPLICATION NUMBER:  08/541,373
39 FILING DATE:  10-OCT-1995
40
41 PRIOR APPLICATION DATA:
42 APPLICATION NUMBER:  08/462,798
43 FILING DATE:  05-JUN-1995
44
45 ATTORNEY/AGENT INFORMATION:
46 NAME:  Carter, Charles G.
47 REGISTRATION NUMBER:  35,093
48 REFERENCE/DOCKET NUMBER:  8648.49USF1
49
50 TELECOMMUNICATION INFORMATION:
51 TELEPHONE:  612/371-5278
52 TELEFAX:  612/332-9081
53
54 TELEX:
55 INFORMATION FOR SEQ ID NO:  37:
56 SEQUENCE CHARACTERISTICS:
57 LENGTH:  372 base pairs
58 TYPE:  nucleic acid
59 STRANDEDNESS:  single
60 TOPOLOGY:  linear
61 MOLECULE TYPE:  Genomic DNA
62 HYPOTHETICAL:  NO
63 ANTI-SENSE:  NO
64 FRAGMENT TYPE:
65 ORIGINAL SOURCE:
66 FEATURE:
67 NAME/KEY:  Coding Sequence
68 LOCATION:  1..372
69 OTHER INFORMATION:
70
71 US-08-767-128-37

```

Query Match	89.38;	Score 13.4;	DB 3;	Length 372;
Best Local Similarity	93.38;	Pred. No. 69;		
Matches	14;	Conservative	0;	Mismatches
			1;	Indels
				Gaps
				0;
QY	1	GACTACTACATGAGC	15	

Db 91 GATTACTACATGAGC 105

RESULT 7

US-08-471-033-18
; Sequence 18, Application US/08471033
; Patent No. 5770696
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,033
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: P-40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-033-18

Query Match 89.3%; Score 13.4; DB 1; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GACTACTACATGAGC 15

Db 1381 GACTACTACATGAGC 1395

RESULT 8

US-08-471-044-18
; Sequence 18, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalin M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2004 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..2004
; OTHER INFORMATION: /note= "Maize optimized DNA
; OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"
US-08-471-044-18

Query Match 89.3%; Score 13.4; DB 2; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
 |||||
 Db 1381 GACTACTACATCAGC 1395

RESULT 9

US-08-463-483A-18
 ; Sequence 18, Application US/08463483A
 ; Patent No. 5849870

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W
 APPLICANT: Koziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: No. 5849870el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463,483A
 FILING DATE:
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Spruill, W. Murray
 REGISTRATION NUMBER: 32,943
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8615
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2004 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1..2004
 OTHER INFORMATION: /note= "Maize optimized DNA
 OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"

US-08-463-483A-18

Query Match 89.3%; Score 13.4; DB 2; Length 2004;
 Best Local Similarity 93.3%; Pred. No. 82;

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
 |||||
 Db 1381 GACTACTACATCAGC 1395

RESULT 10

US-08-471-046A-18
 ; Sequence 18, Application US/08471046A
 ; Patent No. 5866326

GENERAL INFORMATION:

APPLICANT: Warren, Gregory W
 APPLICANT: Koziel, Michael G
 APPLICANT: Mullins, Martha A
 APPLICANT: Nye, Gordon J
 APPLICANT: Carr, Brian
 APPLICANT: Desai, Nalini M
 APPLICANT: Kostichka, N. Kristy
 APPLICANT: Duck, Nicholas B
 APPLICANT: Estruch, Juan J
 TITLE OF INVENTION: Method For Isolating Vegetative Insecticidal
 TITLE OF INVENTION: Protein Genes
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: No. 5866326artis Corporation
 STREET: 3054 Cornwallis Road
 CITY: Research Triangle Park
 STATE: NC
 COUNTRY: USA
 ZIP: 27709

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,046A
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Meigs, J. Timothy
 REGISTRATION NUMBER: 38,241
 REFERENCE/DOCKET NUMBER: CGC1695/CIP3/DIV8 - SOLV4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8587
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2004 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FEATURE:
 NAME/KEY: misc_feature
 LOCATION: 1..2004
 OTHER INFORMATION: /note= "Maize optimized DNA
 OTHER INFORMATION: sequence for VIP1A(a) 80 kd protein from AB78"

US-08-471-046A-18

Query Match 89.3%; Score 13.4; DB 2; Length 2004;
 Best Local Similarity 93.3%; Pred. No. 82;


```

FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
US-08-469-334-18

```

```

Query Match      89.3%; Score 13.4; DB 2; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 GACTACTACATGAGC 15
Db      1381 GACTACTACATCAGC 1395

```

RESULT 13

```

US-09-300-529-18
Sequence 18, Application US/09300529
Patent No. 6066783

```

GENERAL INFORMATION:

```

APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: Genes Encoding Insecticidal Proteins
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 6066783artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: NC
COUNTRY: USA
ZIP: 27709

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/300,529
FILING DATE: TBA
CLASSIFICATION:

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/469,334
FILING DATE: 06-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/463,483
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19506L
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 2004 base pairs

```

```

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..2004
OTHER INFORMATION: /note="Maize optimized DNA
OTHER INFORMATION: sequence for VIPIA(a) 80 kd protein from AB78"
US-09-300-529-18

```

```

Query Match      89.3%; Score 13.4; DB 3; Length 2004;
Best Local Similarity 93.3%; Pred. No. 82;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY      1 GACTACTACATGAGC 15
Db      1381 GACTACTACATCAGC 1395

```

RESULT 14

```

US-08-471-033-35
Sequence 35, Application US/08471033
Patent No. 5770696

```

GENERAL INFORMATION:

```

APPLICANT: Warren, Gregory W
APPLICANT: Kozziel, Michael G
APPLICANT: Mullins, Martha A
APPLICANT: Nye, Gordon J
APPLICANT: Carr, Brian
APPLICANT: Desai, Nalini M
APPLICANT: Kostichka, N. Kristy
APPLICANT: Duck, Nicholas B
APPLICANT: Estruch, Juan J
TITLE OF INVENTION: No. 5770696el Pesticidal Proteins and Strains
NUMBER OF SEQUENCES: 50
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY
COUNTRY: USA
ZIP: 10532

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,033
FILING DATE:
CLASSIFICATION: 530

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/314,594
FILING DATE: 09-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/218,018
FILING DATE: 23-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/037,057
FILING DATE: 25-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Gary M.
REGISTRATION NUMBER: P-40,403
REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SQLv3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8582
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 2576 base pairs

```



```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..2564
; OTHER INFORMATION: /note= "Maize optimized sequence
; OTHER INFORMATION: encoding VIP1A(a) with the Bacillus secretion signal removed
; OTHER INFORMATION: contained in PCIB5526"
US-08-471-033-35

```

```

Query Match      89.3%; Score 13.4; DB 1; Length 2576;
Best Local Similarity 93.3%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACTACTACATGAGC 15
Db      1944 GACTACTACATCAGC 1958

```

```

RESULT 15
US-08-471-044-35
; Sequence 35, Application US/08471044
; Patent No. 5840868
; GENERAL INFORMATION:
; APPLICANT: Warren, Gregory W
; APPLICANT: Kozziel, Michael G
; APPLICANT: Mullins, Martha A
; APPLICANT: Nye, Gordon J
; APPLICANT: Carr, Brian
; APPLICANT: Desai, Nalini M
; APPLICANT: Kostichka, N. Kristy
; APPLICANT: Duck, Nicholas B
; APPLICANT: Estruch, Juan J
; TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,044
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/463,483
; FILING DATE: 05-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/314,594
; FILING DATE: 09-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,018
; FILING DATE: 23-MAR-1994
; APPLICATION NUMBER: US 08/037,057
; FILING DATE: 25-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Gary M.
; REGISTRATION NUMBER: 40,403
; REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8582

```

```

; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2576 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "Synthetic DNA"
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 9..2564
; OTHER INFORMATION: /note= "Maize optimized sequence
; OTHER INFORMATION: encoding VIP1A(a) with the Bacillus secretion signal remov
; OTHER INFORMATION: contained in PCIB5526"
US-08-471-044-35

```

```

Query Match      89.3%; Score 13.4; DB 2; Length 2576;
Best Local Similarity 93.3%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACTACTACATGAGC 15
Db      1944 GACTACTACATCAGC 1958

```

Search completed: June 26, 2003, 04:52:15
Job time : 24.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	393	9	US-09-988-115A-52	Sequence 52, Appl
2	14	93.3	1422	10	US-09-917-800A-1490	Sequence 1490, Ap
3	14	93.3	2000	9	US-09-938-842A-3563	Sequence 3563, Ap
4	14	93.3	2751	10	US-09-935-943-3	Sequence 3, Appli
5	14	93.3	3236	10	US-09-935-943-25	Sequence 25, Appl
6	14	93.3	8002	10	US-09-935-943-10	Sequence 10, Appl
7	13.4	89.3	211	10	US-09-770-696-95	Sequence 95, Appl
8	13.4	89.3	467	10	US-09-864-761-175	Sequence 175, App
9	13.4	89.3	470	9	US-09-918-995-9782	Sequence 9782, Ap
10	13.4	89.3	1197	9	US-10-029-180-63	Sequence 63, Appl
11	13.4	89.3	1571	10	US-09-925-301-93	Sequence 93, Appl
12	13.4	89.3	465237	10	US-09-933-267A-1	Sequence 1, Appli
13	13.4	89.3	3309400	9	US-09-738-626-1	Sequence 1, Appli
14	13	86.7	144	10	US-09-864-761-23136	Sequence 23136, A
15	13	86.7	482	10	US-09-864-761-6421	Sequence 6421, Ap
16	13	86.7	1356	10	US-09-815-242-6624	Sequence 6624, Ap
17	13	86.7	1959	9	US-10-102-524-1862	Sequence 1862, Ap
18	13	86.7	3181	9	US-10-098-841-282	Sequence 282, App
19	13	86.7	3743	9	US-10-102-524-1769	Sequence 1769, Ap

20	13	86.7	5248	9	US-09-822-846-76	Sequence 76, Appl
21	13	86.7	5263	9	US-09-374-046A-19	Sequence 19, Appl
22	13	86.7	5263	9	US-10-102-524-1847	Sequence 1847, Ap
23	13	86.7	11597	10	US-09-070-927A-222	Sequence 222, App
24	12.6	84.0	1531	9	US-10-106-698-2012	Sequence 2012, Ap
25	12.4	82.7	25	9	US-10-098-263B-1584	Sequence 1584, Ap
26	12.4	82.7	91	10	US-09-969-373-152	Sequence 152, App
27	12.4	82.7	179	9	US-09-754-853A-281	Sequence 281, App
28	12.4	82.7	203	10	US-09-878-574-13030	Sequence 13030, A
29	12.4	82.7	220	10	US-09-998-598-2246	Sequence 2246, Ap
30	12.4	82.7	263	9	US-09-954-531-657	Sequence 657, App
31	12.4	82.7	263	10	US-09-969-347-238	Sequence 238, App
32	12.4	82.7	265	10	US-09-878-574-6411	Sequence 6411, Ap
33	12.4	82.7	312	9	US-10-066-543-1572	Sequence 1572, Ap
34	12.4	82.7	393	9	US-09-918-995-35180	Sequence 35180, A
35	12.4	82.7	410	9	US-09-918-995-34814	Sequence 34814, A
36	12.4	82.7	429	7	US-08-781-986A-1087	Sequence 1087, Ap
37	12.4	82.7	442	10	US-09-905-243-42	Sequence 42, Appl
38	12.4	82.7	446	9	US-09-796-692-7562	Sequence 7562, Ap
39	12.4	82.7	446	9	US-10-040-862-7562	Sequence 7562, Ap
40	12.4	82.7	469	10	US-09-864-761-15054	Sequence 15054, A
41	12.4	82.7	494	10	US-09-864-761-3642	Sequence 3642, Ap
42	12.4	82.7	530	10	US-09-879-536-682	Sequence 682, App
43	12.4	82.7	587	10	US-09-903-456-16	Sequence 16, Appl
44	12.4	82.7	728	10	US-09-910-943-694	Sequence 694, App
45	12.4	82.7	858	10	US-09-764-877-850	Sequence 850, App

ALIGNMENTS

RESULT 1
US-09-988-115A-52
; Sequence 52, Application US/09988115A
; Publication No. US20030037347A1
; GENERAL INFORMATION:
; APPLICANT: Robl, James M.
; APPLICANT: Goldsby, Richard A.
; APPLICANT: Ferguson, Stacy E.
; APPLICANT: Kuroiwa, Yoshima
; APPLICANT: Tomizuka, Kazuma
; APPLICANT: Ishida, Isao
; TITLE OF INVENTION: Expression of Xenogenous (Human)
; TITLE OF INVENTION: Immunoglobulins in Cloned, Transgenic Ungulates
; FILE REFERENCE: 50195/008003
; CURRENT APPLICATION NUMBER: US/09/988, 115A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,625
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 60/256,458
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 09/714,185
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: US 60/166,410
; PRIOR FILING DATE: 1999-11-19
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 393
; TYPE: DNA
; ORGANISM: Bovine
; US-09-988-115A-52

Query Match 100.0%; Score 15; DB 9; Length 393;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
Db 67 GACTACTACATGAGC 81

RESULT 2

```

US-09-917-800A-1490
; Sequence 1490, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castile, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1490
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 X13016
US-09-917-800A-1490

Query Match          93.3%; Score 14; DB 10; Length 1422;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAG 14
        |||||||||
Db      347 GACTACTACATGAG 360

RESULT 3
US-09-938-842A-3563/C
; Sequence 3563, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 3563

```

```

; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-3563

Query Match          93.3%; Score 14; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ACTACTACATGAGC 15
        |||||||||
Db      1589 ACTACTACATGAGC 1576

RESULT 4
US-09-935-943-3
; Sequence 3, Application US/09935943
; Patent No. US20020120963A1
; GENERAL INFORMATION:
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Wegrich, Lynette M.
; APPLICANT: Budziszewski, Gregory J.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-31509A
; CURRENT APPLICATION NUMBER: US/09/935,943
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: PCT/EP01/08910
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US 60/222,779
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 2751
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2751)
US-09-935-943-3

Query Match          93.3%; Score 14; DB 10; Length 2751;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 ACTACTACATGAGC 15
        |||||||||
Db      1998 ACTACTACATGAGC 2011

RESULT 5
US-09-935-943-25
; Sequence 25, Application US/09935943
; Patent No. US20020120963A1
; GENERAL INFORMATION:
; APPLICANT: Levin, Joshua Z.
; APPLICANT: Wegrich, Lynette M.
; APPLICANT: Budziszewski, Gregory J.
; TITLE OF INVENTION: Herbicide Target Genes and Methods
; FILE REFERENCE: PB/5-31509A
; CURRENT APPLICATION NUMBER: US/09/935,943
; CURRENT FILING DATE: 2001-08-23
; PRIOR APPLICATION NUMBER: PCT/EP01/08910
; PRIOR FILING DATE: 2001-08-01
; PRIOR APPLICATION NUMBER: US 60/222,779
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 3236
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:

```

NAME/KEY: CDS
LOCATION: (1)..(1056)
US-09-935-943-25

Query Match 93.3%; Score 14; DB 10; Length 3236;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
|||||
Db 2295 ACTACTACATGAGC 2308

RESULT 6
US-09-935-943-10

Sequence 10, Application US/09935943
Patent No. US20020120963A1
GENERAL INFORMATION:
APPLICANT: Levin, Joshua Z.
APPLICANT: Wegrich, Lynette M.
APPLICANT: Budziszewski, Gregory J.
TITLE OF INVENTION: Herbicide Target Genes and Methods
FILE REFERENCE: PB/5-31509A
CURRENT APPLICATION NUMBER: US/09/935,943
PRIOR FILING DATE: 2001-08-23
PRIOR APPLICATION NUMBER: PCT/EP01/08910
PRIOR FILING DATE: 2001-08-01
PRIOR APPLICATION NUMBER: US 60/222,779
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 26
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 8002
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-935-943-10

Query Match 93.3%; Score 14; DB 10; Length 8002;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
|||||
Db 6239 ACTACTACATGAGC 6252

RESULT 7
US-09-770-696-95/c

Sequence 95, Application US/09770696
Patent No. US20010044940A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krickler, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
TITLE OF INVENTION: thaliana
FILE REFERENCE: 2031US (PARA-020PRV)

CURRENT APPLICATION NUMBER: US/09/770,696
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/178,278
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 911
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 95
LENGTH: 211
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-696-95

Query Match 89.3%; Score 13.4; DB 10; Length 211;
Best Local Similarity 93.3%; Pred. No. 3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15
|||||
Db 148 GACTACACATGAGC 134

RESULT 8
US-09-864-761-175

Sequence 175, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1

```

; SEQ ID NO 175
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034402.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 36
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5e+02
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 34
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 44
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2e+02
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3e+02
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.3e+02
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 46
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 68
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 77
US-09-864-761-175

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 467;
Best Local Similarity 93.3%; Pred. No. 3.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACTACTACATGAGC 15
        ||||| ||||| |||||
Db      184 GACTCCTACATGAGC 198

```

```

RESULT 9
US-09-918-995-9782
; Sequence 9782, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918, 995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235, 076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9782
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(470)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-9782

```

```

Query Match      89.3%; Score 13.4; DB 9; Length 470;
Best Local Similarity 93.3%; Pred. No. 3.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACTACTACATGAGC 15
        ||||| ||||| |||||
Db      104 GACTACACATGAGC 118

```

```

RESULT 10
US-10-029-180-63
; Sequence 63, Application US/10029180
; Publication No. US20020182708A1
; GENERAL INFORMATION:
; APPLICANT: Cali, Brian M.
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin T.
; APPLICANT: Milna, G. Todd
; APPLICANT: Sherman, Amlr
; APPLICANT: Silva, Jeffrey C.
; APPLICANT: Trueheart, Josh

```

```

; APPLICANT: Zhang, Lixin
; TITLE OF INVENTION: NO. US20020182708A1e1 Regulators of Fungal Gene Expression
; FILE REFERENCE: MIC-004
; CURRENT APPLICATION NUMBER: US/10/029,180
; CURRENT FILING DATE: 2001-12-22
; PRIOR APPLICATION NUMBER: US 60/257,431
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 138
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 63
; LENGTH: 1197
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: fungal gene
US-10-029-180-63

```

```

Query Match      89.3%; Score 13.4; DB 9; Length 1197;
Best Local Similarity 93.3%; Pred. No. 3.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACTACTACATGAGC 15
        ||||| ||||| |||||
Db      754 GACTACTACATGAGC 768

```

```

RESULT 11
US-09-925-301-93/c
; Sequence 93, Application US/09925301
; Patent No. US20020052308A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 1571
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1497)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1516)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1530)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1546)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1571)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-93

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 1571;
Best Local Similarity 93.3%; Pred. No. 3.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 GACTACTACATGAGC 15
        ||||| ||||| |||||
Db      899 GACTACTCCTACATGAGC 885

```



```

; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
; OTHER INFORMATION: SWISSPROT HIT: P51957, EVALUE 4.00e-23
; OTHER INFORMATION: EST_HUMAN HIT: AUI25616.1, EVALUE 3.00e-73
; OTHER INFORMATION: NT HIT: g111435215, EVALUE 8.00e-76
US-09-864-761-23136

```

```

Query Match      86.7%; Score 13; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 CTACTACATGAGC 15
        |||
Db      96 CTACTACATGAGC 108

```

```

RESULT 15
US-09-864-761-6421
; Sequence 6421, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6421
; LENGTH: 482
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006254.10

```

```

; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.3
US-09-864-761-6421

```

```

Query Match      86.7%; Score 13; DB 10; Length 482;
Best Local Similarity 100.0%; Pred. No. 5.5e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3 CTACTACATGAGC 15
        |||
Db      420 CTACTACATGAGC 432

```

```

Search completed: June 26, 2003, 04:56:59
Job time : 68.2 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-13
Perfect score: 15
Sequence: 1 gactactacatgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1:	em_estba:*
2:	em_esthum:*
3:	em_estin:*
4:	em_estmu:*
5:	em_estov:*
6:	em_estpl:*
7:	em_estro:*
8:	em_hlc:*
9:	gb_estl:*
10:	gb_est2:*
11:	gb_hlc:*
12:	gb_est3:*
13:	gb_est4:*
14:	gb_est5:*
15:	em_estfun:*
16:	em_estom:*
17:	gb_gss:*
18:	em_gss_hum:*
19:	em_gss_inv:*
20:	em_gss_pln:*
21:	em_gss_vrt:*
22:	em_gss_fun:*
23:	em_gss_mam:*
24:	em_gss_mus:*
25:	em_gss_other:*
26:	em_gss_pro:*
27:	em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	308	10	AW801084	AW801084 MR3-UM006
2	15	100.0	487	10	AW403862	AW403862 UI-HF-BK0
3	15	100.0	669	12	BG686716	BG686716 602650678
4	15	100.0	942	14	BQ710859	BQ710859 AGENCOURT
5	15	100.0	1112	12	BG760090	BG760090 602733346
6	15	100.0	1164	12	BF974771	BF974771 602245420

C	7	14	93.3	230	9	AV272811	AV272811
C	8	14	93.3	250	9	AJ237250	AJ237250
C	9	14	93.3	251	9	AV272672	AV272672
C	10	14	93.3	360	14	C50628	C50628
C	11	14	93.3	377	14	C50318	C50318
C	12	14	93.3	405	12	BF919229	BF919229
C	13	14	93.3	410	17	AQ019173	AQ019173
C	14	14	93.3	417	12	BF816356	BF816356
C	15	14	93.3	420	9	AA702148	AA702148
C	16	14	93.3	442	17	BH522357	BH522357
C	17	14	93.3	451	17	BH565724	BH565724
C	18	14	93.3	455	9	AL366023	AL366023
C	19	14	93.3	463	9	AL371592	AL371592
C	20	14	93.3	483	17	AQ133642	AQ133642
C	21	14	93.3	498	10	AW342506	AW342506
C	22	14	93.3	499	13	BM525120	BM525120
C	23	14	93.3	512	13	BM524599	BM524599
C	24	14	93.3	514	17	AQ396842	AQ396842
C	25	14	93.3	518	12	BG664310	BG664310
C	26	14	93.3	526	9	AL370601	AL370601
C	27	14	93.3	537	12	BF289546	BF289546
C	28	14	93.3	543	12	BG664580	BG664580
C	29	14	93.3	554	13	BG958385	BG958385
C	30	14	93.3	556	12	BG665152	BG665152
C	31	14	93.3	590	9	AI855171	AI855171
C	32	14	93.3	595	10	BE425840	BE425840
C	33	14	93.3	631	12	BG669077	BG669077
C	34	14	93.3	637	10	BE425828	BE425828
C	35	14	93.3	649	13	BJ234400	BJ234400
C	36	14	93.3	661	11	AY067288	AY067288
C	37	14	93.3	661	14	BQ507135	BQ507135
C	38	14	93.3	677	12	BG597124	BG597124
C	39	14	93.3	688	13	BJ305740	BJ305740
C	40	14	93.3	742	10	BE383112	BE383112
C	41	14	93.3	746	14	BQ804288	BQ804288
C	42	14	93.3	773	14	BQ804369	BQ804369
C	43	14	93.3	814	17	AG050744	AG050744
C	44	14	93.3	835	14	BQ507136	BQ507136
C	45	14	93.3	857	12	BF693325	BF693325

ALIGNMENTS

RESULT 1
AW801084/c 308 bp mRNA linear EST 16-MAY-2000
LOCUS MR3-UM0065-030300-107-a09 UM0065 Homo sapiens cDNA, mRNA sequence.
DEFINITION AW801084
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
human.
Homo sapiens

REFERENCE
AUTHORS
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 308)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL
MEDLINE
COMMENT
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2=MR3-UM0065-030
300-107-a09&tl3=2000-03-03&tl4=1)
Seq primer: puc 18 forward
High quality sequence stop: 308.
Location/Qualifiers

FEATURES

source

1. 308

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="UM0065"

/dev_stage="Adult"

/note="Organ: uterus; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 65 a 89 c 81 g 73 t
ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 308;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15

Db 190 GACTACTACATGAGC 176

RESULT 2

LOCUS

AM403862 487 bp mRNA linear EST 16-FEB-2000

DEFINITION UI-HF-BK0-abo-a-12-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone

IMAGE:3056926 5', mRNA sequence.

AM403862

AM403862.1 GI:6923015

EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 487)

NIH-MGC http://mgs.nci.nih.gov/.

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Eco RI site shown at the beginning of the sequence.

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: M.B. Soares Lab

CDNA Library Arrayed by: M.B. Soares Lab

DNA Sequencing by: M.B. Soares Lab

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Seq primer: M13 Forward.

Location/Qualifiers

1. 487

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:3056926"

/clone_lib="NIH_MGC_36"

/tissue_type="lymph"

/cell_line="germinal center B cells"

/lab_host="DH10B (LTI)"

/note="Vector: pT7T3-pac; Site_1: NotI; Site_2: Eco RI;

Constructed from size fractionated cytoplasmic mRNA

(0.5-1.5kb). Directionally cloned. Cells provided by Louis

M. Staudt, Ph.D. Library preparation by Maria de Fatima

BASE COUNT 109 a 115 c 143 g 120 t
ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 487;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15

Db 215 GACTACTACATGAGC 229

RESULT 3

LOCUS

BG686716 669 bp mRNA linear EST 01-MAY-2001

DEFINITION 602650678F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4763274 5',

mRNA sequence.

BG686716

BG686716.1 GI:13918113

EST.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 669)

NIH-MGC http://mgs.nci.nih.gov/.

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCM1618 row: k column: 19

High quality sequence stop: 665.

Location/Qualifiers

1. 669

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4763274"

/clone_lib="NIH_MGC_48"

/tissue_type="primary B-cells from tonsils (cell line)"

/lab_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected >500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT 149 a 184 c 186 g 150 t

ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 669;

Best Local Similarity 100.0%; Pred. No. 1.3e+03;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAGC 15

Db 217 GACTACTACATGAGC 231

RESULT 4 942 bp mRNA linear EST 16-JUL-2002

LOCUS BQ710859

DEFINITION AGENCOURT_8501500 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6303061

ACCESSION	5	5', mRNA sequence.
VERSION	B0710859	
KEYWORDS	B0710859.1	GI:21849758
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 942)	
TITLE	NIH-MGC http://mgc.nci.nih.gov/.	
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)	
COMMENT	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgabs-remail.nih.gov	
	Tissue Procurement: Dr. Mark Watson	
	CDNA Library Preparation: Rubin Laboratory	
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
	DNA Sequencing by: Agencourt Bioscience Corporation	
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:	
	http://image.llnl.gov	
	Plate: LLCM2521	row: 1 column: 14
	High quality sequence stop: 510.	
FEATURES	Location/Qualifiers	
source	1..942	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:6303061"	
	/clone_lib="NIH_MGC_113"	
	/lab_host="DH10B (phage-resistant)"	
	/note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."	
BASE COUNT	197 a	291 c 263 g 191 t
ORIGIN		
Query Match	100.0%;	Score 15; DB 14; Length 942;
Best Local Similarity	100.0%;	Pred. No. 1.5e+03;
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
OY	1	GACTACTACATGAGC 15
Db	208	GACTACTACATGAGC 222
RESULT 5		
LOCUS	BG760090	
DEFINITION	BG760090	1112 bp mRNA linear EST 15-MAY-2001
ACCESSION	602733346F1	NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4878851 5',
VERSION	BG760090	mRNA sequence.
KEYWORDS	BG760090.1	GI:14070743
SOURCE	EST.	
ORGANISM	human.	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
	1 (bases 1 to 1112)	
	NIH-MGC http://mgc.nci.nih.gov/.	
	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	
	Contact: Robert Strausberg, Ph.D.	
	Email: cgabs-remail.nih.gov	
	Tissue Procurement: Louis M. Staudt, M.D., Ph.D.	
	CDNA Library Preparation: Ling Hong/Rubin Laboratory	
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)	
	DNA Sequencing by: Incyte Genomics, Inc.	
	Clone distribution: MGC clone distribution information can be	

```

found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1765 row: k column: 12
High quality sequence stop: 509.
Location/Qualifiers
1. 1112
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4878851"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library."

BASE COUNT      346 a      268 c      318 g      180 t
ORIGIN

Query Match
Best Local Similarity 100.0%; Score 15; DB 12; Length 1112;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 GACTACTACATGAGC 15
        |||
Db      218 GACTACTACATGAGC 232

RESULT 6
BF974771      1164 bp      mRNA      linear      EST 22-JAN-2001
LOCUS      602245420F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4336541 5',
DEFINITION      mRNA sequence.
ACCESSION      BF974771
VERSION      BF974771.1 GI:12341986
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1164)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1995)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1209 row: g column: 06
High quality sequence stop: 696.
Location/Qualifiers
1. 1164
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4336541"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH-MGC library."

```


Db 157 GACTACTACATGAG 170

|||||

RESULT 9
LOCUS AV272672/c 251 bp mRNA linear EST 05-NOV-1999
DEFINITION AV272672 RIKEN full-length enriched, adult male testis (DH10B) Mus musculus cDNA clone 4931431F16 3', mRNA sequence.

ACCESSION AV272672
VERSION AV272672
KEYWORDS AV272672.1 GI:6260709
SOURCE EST.
ORGANISM house mouse.
Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 251)
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomioka,N., Tsunoda,Y., Watanabe,H., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Konno,H., et al. 1999)

TITLE Unpublished (1999)
JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
COMMENT The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gscc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki,N., Izawa,M., Watanabe,H., Okazaki,Y. and Hayashizaki,Y.
Matsuyama,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh,M., Katsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
source Location/Qualifiers
1. 251
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4931431F16"
/clone_lib="RIKEN full-length enriched, adult male testis (DH10B)"
/sex="male"
/tissue_type="testis"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site_1: SalI; Site_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5',
GAGAGAGAGAGATCCACAGCTCTTTT TTTT TTTT TTTT VN 3'], cDNA was

prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5',
GAGAGAGAGATCTCGAGTTAATTAATTAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 75 a 41 c 49 g 86 t
ORIGIN

Query Match 93.3%; Score 14; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
|||||

Db 199 ACTACTACATGAGC 186

RESULT 10
LOCUS C50628 360 bp mRNA linear EST 18-OCT-1999
DEFINITION C50628 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
Caenorhabditis elegans cDNA clone yk497b1 5', mRNA sequence.

ACCESSION C50628
VERSION C50628.1 GI:2387881
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 360)
Kohara,Y., Mutohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano,M., Miyata,A. and Nishigaki,A.
Expression map of the C.elegans genome
Unpublished (1996)
Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

TITLE Unpublished (1996)
JOURNAL Expression map of the C.elegans genome
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source Location/Qualifiers
1. 360
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk497b1"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"

BASE COUNT 105 a 74 c 81 g 98 t 2 others
ORIGIN

Query Match 93.3%; Score 14; DB 14; Length 360;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
|||||

Db 105 ACTACTACATGAGC 118

RESULT 11
LOCUS C50318 377 bp mRNA linear EST 18-OCT-1999
DEFINITION C50318 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
Caenorhabditis elegans cDNA clone yk480c3 5', mRNA sequence.

ACCESSION C50318
VERSION C50318.1 GI:2387571

KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 377)
AUTHORS Kohara,Y., Motohashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.

FEATURES
source
1.377
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk480c3"
/clone_lib="Yuji Kohara unpublished cDNA:strain N2
hermaphrodite embryo"
/sex="hermaphrodite"
/dev_stage="embryo"
BASE COUNT 112 a 76 c 86 g 102 t 1 others
ORIGIN

Query Match 93.3%; Score 14; DB 14; Length 377;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ACTACTACATGAGC 15
|||||
Db 114 ACTACTACATGAGC 127

RESULT 12
BF919229 405 bp mRNA linear EST 19-JAN-2001
LOCUS CMO-NT0131-281000-641-g10 NT0131 Homo sapiens cDNA, mRNA sequence.
DEFINITION BF919229
ACCESSION BF919229
VERSION BF919229.1 GI:12315117
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 405)
REFERENCE Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CMO&tt2=CMO-NT0131-
281000-641-g10&tt3=2000-10-28&tt4=1)

Seq primer: puc 18 forward
High quality sequence start: 9
High quality sequence stop: 405.
location/Qualifiers
1.405
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NT0131"
/dev_stage="Adult"
/note="Organ: nervous_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. letters patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 125 a 88 c 57 g 135 t
ORIGIN

Query Match 93.3%; Score 14; DB 12; Length 405;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 ACTACTACATGAGC 15
|||||
Db 258 ACTACTACATGAGC 271

RESULT 13
AQ019173 410 bp DNA linear GSS 09-JUN-1998
LOCUS CIT-HSP-2303C19.TR CIT-HSP Homo sapiens genomic clone 2303C19, DNA
DEFINITION sequence.
ACCESSION AQ019173
VERSION AQ019173.1 GI:3197909
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 410)
REFERENCE Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2303C19.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.
location/Qualifiers
1.410
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2303C19"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 123 a 91 c 56 g 140 t
ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 410;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
|||||
Db 241 ACTACTACATGAGC 254

RESULT 14
BF816356 417 bp mRNA linear EST 13-JAN-2001
DEFINITION MR3-CI0185-061200-002-b11 CI0185 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF816356
VERSION BF816356.1 GI:12151918
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 417)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR3&tl2=MR3-CI0185-
061200-002-b11&tl3=2000-12-06&tl4=1)
Seq primer: puc 18 forward
High quality sequence stop: 416.

FEATURES
source
1. 417
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CI0185"
/dev_stage="Adult"
/note="Organ: colon.ins; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 85 a 109 c 143 g 80 t
ORIGIN

Query Match 93.3%; Score 14; DB 12; Length 417;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTACTACATGAG 14
|||||
Db 8 GACTACTACATGAG 21

RESULT 15
AA702148/c 420 bp mRNA linear EST 19-DEC-1997
LOCUS AA702148

DEFINITION z185h12.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
clone IMAGE:447623 3' similar to gb:X59543_final
RIBONUCLEOSIDE-DIPHOSPHATE REDUCTASE M1 CHAIN (HUMAN);, mRNA
sequence.
ACCESSION AA702148
VERSION AA702148.1 GI:2705261
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 420)
Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
White,Y., Wylie,T., Waterston,R. and Wilson,R.
WashU-NCI human EST Project
Unpublished (1997)
TITLE WashU-NCI human EST Project
JOURNAL
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Trace considered overall poor quality
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 1.

FEATURES
source
1. 420
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="GDB:1351880"
/db_xref="taxon:9606"
/clone="IMAGE:447623"
/clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a substracted version of the original Soares fetal
liver spleen 1NFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5',
AACTGGAAGATTATTAATTAAGATCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 109 a 105 c 70 g 136 t
ORIGIN

Query Match 93.3%; Score 14; DB 9; Length 420;
Best Local Similarity 100.0%; Pred. No. 3.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTACTACATGAGC 15
|||||
Db 176 ACTACTACATGAGC 163

Search completed: June 26, 2003, 04:50:09
Job time : 753.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-25

Perfect score: 15

Sequence: 1 acctaccgcatgcac 15

Scoring table: IDENTITY_NUC

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: gb_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	6	ARI83486	ARI83486 Sequence
2	15	100.0	370	9	AF029333	AF029333 Homo sapi
3	15	100.0	717	9	AF048775	AF048775 Homo sapi
4	15	100.0	2030	9	AK090859	AK090859 Homo sapi
5	15	100.0	42845	9	HSU73633	U73633 Human chrom
6	15	100.0	114287	9	AP000590	AP000590 Homo sapi
7	15	100.0	157226	9	AC018902	AC018902 Homo sapi
8	15	100.0	194201	2	AP000896	AP000896 Homo sapi
9	15	100.0	210402	9	AP001271	AP001271 Homo sapi
10	15	100.0	214415	9	AC010674	AC010674 Homo sapi
11	15	100.0	333800	1	SME591792	AL591792 Sinorhizo
12	14	93.3	787	9	HSIGGVDJ2	X81696 H.sapiens r
13	14	93.3	38891	2	AC010298	AC010298 Homo sapi
14	14	93.3	42432	1	AF083072	AF083072 Cenarchae
15	14	93.3	72354	9	AL357123	AL357123 Human DNA
16	14	93.3	83629	2	AC110485	AC110485 Rattus no
17	14	93.3	110157	2	AC130966	AC130966 Rattus no
18	14	93.3	151438	2	AC103516	AC103516 Rattus no
19	14	93.3	152835	8	AP004224	AP004224 Oryza sat
20	14	93.3	154681	2	AC022000	AC022000 Homo sapi
21	14	93.3	159677	2	AC113802	AC113802 Rattus no
22	14	93.3	161722	2	AC121969	AC121969 Mus muscu
23	14	93.3	162742	2	AC096676	AC096676 Homo sapi
24	14	93.3	164764	2	AC102957	AC102957 Rattus no
25	14	93.3	176542	2	AC097249	AC097249 Rattus no
26	14	93.3	221380	2	AC094347	AC094347 Rattus no
27	14	93.3	341720	2	AC126939	AC126939 Mus muscu
28	13.4	89.3	231	9	HSA407974	AJ407974 Homo sapi
29	13.4	89.3	242	9	HSA402495	AJ402495 Homo sapi
30	13.4	89.3	273	9	HSJ8229	AJ008229 Homo sapi
31	13.4	89.3	288	9	HSZ80501	Z80501 H.sapiens B
32	13.4	89.3	290	9	AF077438	AF077438 Homo sapi
33	13.4	89.3	305	9	HSZ80655	Z80655 H.sapiens B
34	13.4	89.3	308	9	AY003764	AY003764 Homo sapi
35	13.4	89.3	311	9	AY003712	AY003712 Homo sapi
36	13.4	89.3	329	9	AY003723	AY003723 Homo sapi
37	13.4	89.3	329	9	AY003765	AY003765 Homo sapi
38	13.4	89.3	346	9	HSIGHVBAH	Z18836 Homo sapien
39	13.4	89.3	354	9	HSA132850	AJ132850 Homo sapi
40	13.4	89.3	354	9	HSA271549	AJ271549 Homo sapi
41	13.4	89.3	357	9	HSA132851	AJ132851 Homo sapi
42	13.4	89.3	363	9	HSA245298	AJ245298 Homo sapi
43	13.4	89.3	363	9	HSA245299	AJ245299 Homo sapi
44	13.4	89.3	363	9	HSA245300	AJ245300 Homo sapi
45	13.4	89.3	372	6	ARI60944	ARI60944 Sequence

ALIGNMENTS

RESULT 1
ARI83486
LOCUS ARI83486 15 bp DNA
DEFINITION Sequence 25 from patent US 6342220.
ACCESSION ARI83486
VERSION ARI83486.1 GI:20227455
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 25 29-JAN-2002;
FEATURES Location/Qualifiers

source	1. .15		
BASE COUNT	4 a	6 c	3 g
ORIGIN	2 t		
Query Match	100.0%;	Score 15;	DB 6;
Best Local Similarity	100.0%;	Pred. No. 2.5e+04;	Length 15;
Matches	15;	Conservative 0;	Mismatches 0;
		Indels 0;	Gaps 0;
QY	1 ACCTACGGCATGCAC 15		
Db	1 ACCTACGGCATGCAC 15		
RESULT 2			
AF029333	370 bp	mRNA	linear
LOCUS	PRI 10-NOV-1997		
DEFINITION	Homo sapiens puuA1 immunoglobulin heavy chain V region mRNA,		
ACCESSION	AF029333		
VERSION	AF029333.1	GI:2599529	
KEYWORDS	.		
SOURCE	Homo sapiens.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheraia; Primates; Catarrhini; Homidae; Homo.		
TITLE	1 (bases 1 to 370)		
JOURNAL	Salonen,E.-M., Parren,P.W.H.I., Graus,Y.F., Lundkvist,A.,		
REFERENCE	Fisicaro,P., Vapalahti,O., Kallio-Kokko,H., Vaheri,A. and		
AUTHORS	Burton,D.R.		
TITLE	Human recombinant Puumala virus antibodies: cross reaction with		
JOURNAL	other hantaviruses and use in diagnostics		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 370)		
TITLE	Salonen,E.-M., Parren,P.W.H.I., Graus,Y.F., Lundkvist,A.,		
JOURNAL	Fisicaro,P., Vapalahti,O., Kallio-Kokko,H., Vaheri,A. and		
REFERENCE	Burton,D.R.		
AUTHORS	Direct Submission		
TITLE	Submitted (09-OCT-1997) Immunology, The Scripps research Institute,		
JOURNAL	10550 N. Torrey Pines Road, La Jolla, CA 92037, USA		
FEATURES	Location/Qualifiers		
source	1. 370		
	/organism="Homo sapiens"		
	/isolate="puuA1"		
	/db_xref="taxon:9606"		
	<1. >370		
	/note="Fd fragment; recognizes Puumala virus G2		
	glycoprotein (G2c epitope)"		
	/codon_start=1		
	/product="immunoglobulin heavy chain V region"		
	/protein_id="AAB84156.1"		
	/db_xref="GI:2599530"		
	/translation="LLESGGVVQPGRSRLSCVASGFTFTYGMHWVROAPGKLEW		
	VATISVDGSOKDYADSLKGRFTISRDNSTVYLLQMSLRADTAVYICAKAQENYDT		
	RGAIVHWGGSIVVSSASTK"		
BASE COUNT	84 a	102 c	103 g
ORIGIN	81 t		
Query Match	100.0%;	Score 15;	DB 9;
Best Local Similarity	100.0%;	Pred. No. 8.5e+03;	Length 370;
Matches	15;	Conservative 0;	Mismatches 0;
		Indels 0;	Gaps 0;
QY	1 ACCTACGGCATGCAC 15		
Db	82 ACCTACGGCATGCAC 96		
RESULT 3			
AF048775	717 bp	mRNA	linear
LOCUS	PRI 31-DEC-1998		
DEFINITION	Homo sapiens clone 12B5 anti-Mpl scFv mRNA, partial cds.		
ACCESSION	AF048775		
VERSION	AF048775.1	GI:2911501	

KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 717)
AUTHORS Merchant,A.M., Zhu,Z., Yuan,J.Q., Goddard,A., Adams,C.W., Presta,L.G. and Carter,P.
An efficient route to human bispecific IgG
Nat. Biotechnol. 16 (7), 677-681 (1998)

TITLE 98325681
JOURNAL PUBMED 9661204
REFERENCE 2 (bases 1 to 717)
AUTHORS Goddard,A., Yuan,J.Q., Zhu,Z. and Carter,P.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1998) Molecular Oncology, Genentech Inc, 1 DNA Way, South San Francisco, CA 94080, USA

FEATURES
source Location/Qualifiers
1..717
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="12B5"
<1..>717
/codon_start=1
/product="anti-Mpl scFv"
/protein_id="AAC98736.1"
/db_xref="GI:2911502"
/translation="QVQLVQSGGGLYVRPGSLSLSCAVSGITLRTYGMHWVRQAPGKGLEWVAGISFDGSEYADSVQGRFTISRSSKNTLYLQMNSLRADYAVYCARGAHYGFDIWQGDTMYTVVSSGGGTGGGGSGGSDIQMTQSPSTLSASIGDRYITCRASEGTYHWLAWYQOKPKGAPKLLIYKASSLSASCAPSRFSGSGSDFTLTISSLQPDGFATY YCQQSYNYPLEFGGCTKLEIK"

BASE COUNT 161 a 188 c 215 g 153 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 717;
Best Local Similarity 100.0%; Pred. No. 6.8e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
db 91 ACCTACGGCATGCAC 105

RESULT 4
AK090859/c
LOCUS 2030 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ33540 fis, clone BRAMY2007613.
ACCESSION AK090859
VERSION AK090859.1 GI:21749098
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens amygdala cDNA to mRNA, clone_lib:BRAMY2 clone:BRAMY2007613.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S., Fukuzumi,Y., Fujimori,Y., Komiyaama,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Magatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuhara,Y., Nagai,K. and Isoigai,T.
TITLE MEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2030)
AUTHORS Isoigai,T. and Yamamoto,J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isoigai, FLJ Project(HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
source
1. 2030
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="BRAMY2007613"
/tissue_type="amygdala"
/clone_lib="BRAMY2"
/note="cloning vector: PME18SFL3"

BASE COUNT 586 a 345 c 402 g 697 t

ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 2030;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
Db 361 ACCTACGGCATGCAC 347

RESULT 5
HSU73633/c 42845 bp DNA linear PRI 19-JUN-1997
LOCUS Human chromosome 11 146h12 cosmid, complete sequence.
DEFINITION U73633
ACCESSION U73633.1 GI:2199506
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 42845)
Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J., Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L., Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hanher,L., Harris,J., Hinson,S., McFarland,J., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K., Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-1996) McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT On Jun 19, 1997 this sequence version replaced gi:1737188.
FEATURES
source
1. 42845
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q23"
/clone="146h12"
/clone_lib="library prepared from flow sorted human chromosome 11 derived from Chinese hamster ovary (CHO) monochromosomal somatic cell hybrid, J1; vector: scos-1"
2218. 2488
/rpt_family="ALU"
/rpt_type=dispersed
repeat_region complement(12155. 12233)
/rpt_family="MIR"
/rpt_type=dispersed
12594. 12879
/rpt_family="ALU"
/rpt_type=dispersed
13713. 14193
/rpt_family="ALU"

repeat_region /rpt_type=dispersed
14374. 14661
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(14849. 15071)
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
16557. 16630
/rpt_family="MIR"
repeat_region /rpt_type=dispersed
20477. 20764
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(22420. 22688)
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(24053. 24356)
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
26266. 26540
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(27183. 27512)
/rpt_family="MLT1"
repeat_region /rpt_type=dispersed
complement(28625. 28911)
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(32586. 32629)
/rpt_family="L1"
repeat_region /rpt_type=dispersed
complement(32743. 33041)
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(33028. 33172)
/rpt_family="L1"
repeat_region /rpt_type=dispersed
35218. 35473
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(36781. 37041)
/rpt_family="MLT1"
repeat_region /rpt_type=dispersed
39388. 39646
/rpt_family="ALU"
repeat_region /rpt_type=dispersed
complement(39724. 39766)
/rpt_family="MER46"
repeat_region /rpt_type=dispersed
complement(39985. 40055)
/rpt_family="MER46"
/rpt_type=dispersed

BASE COUNT 9620 a 11179 c 11246 g 10800 t

ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 42845;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
Db 1464 ACCTACGGCATGCAC 1450

RESULT 6
AP000590 114287 bp DNA linear PRI 18-JUL-2001
LOCUS AP000590
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:CMB9-52H4, complete sequence.
ACCESSION AP000590.6 GI:14861095
VERSION AP000590.6 HTG.
KEYWORDS Homo sapiens DNA, clone:CMB9-52H4.
SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
REFERENCE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
AUTHORS Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens genomic DNA
JOURNAL Published Only in Database (1999)
REFERENCE 2 (bases 1 to 114287)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
On Jul 17, 2001 this sequence version replaced gi:10880434.
FEATURES
source
1.114287
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="CMB9-52H4"
BASE COUNT 28585 a 30994 c 28832 g 25876 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 114287;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGGCATGCAC 15
Db 22215 ACCTACGGCATGCAC 22229
RESULT 7
AC018902 157226 bp DNA linear PRI 09-AUG-2001
LOCUS Homo sapiens chromosome 15 clone RP11-327C2 map 15Q21.2, complete
DEFINITION sequence.
AC018902
AC018902.8 GI:15138554
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 157226)
REFERENCE Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
AUTHORS Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 157226)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 157226)
REFERENCE Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
AUTHORS Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
On Aug 9, 2001 this sequence version replaced gi:13173608.
COMMENT ----- Genome Center

Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_ww
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note : Data from overlapping BACs AC010674 [drafting center:
UWMSC], AC025917 [drafting center: UWMSC] and AC073438 [drafting
center: WIBR] were added for finishing

FEATURES
source
1.157226
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15Q21.2"
/clone="RP11-327C2"
/clone_lib="RPCT human BAC library 11"
/note="Data from overlapping BACs RP11-430B1, RP11-23N2,
and RP11-472P3 were added and the consensus sequence was
determined from RP11-327C2 to the extent possible"
1.37195
/note="overlap with RP11-430B1 AC010674"
80042.157226
/note="overlap with RP11-472P3 AC073438"
122504.122615
/note="low quality data"
148580.157226
/note="overlap with RP11-23N2 AC025917"
BASE COUNT 49939 a 31674 c 30429 g 45184 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 157226;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGGCATGCAC 15
Db 1953 ACCTACGGCATGCAC 1967
RESULT 8
AP000896 194201 bp DNA linear HTG 30-MAY-2000
LOCUS Homo sapiens chromosome 11 clone RP11-677D19 map 11q13, WORKING
DEFINITION DRAFT SEQUENCE, 61 unordered pieces.
AP000896
AP000896.2 GI:8119039
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-677D19.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 194201)
REFERENCE Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
AUTHORS Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Published Only in Database (1999)
JOURNAL Published Only in Database of 11q13
REFERENCE 2 (bases 1 to 194201)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1999) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
Japan (E-mail:hattori@psc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On May 31, 2000 this sequence version replaced gi:6997719.

----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: <http://hgp.gsc.riken.go.jp/>
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: HumDraft11
Center clone name: RP11-677D19
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 170176 bases at least Q40
Consensus quality: 175910 bases at least Q30
Consensus quality: 181676 bases at least Q20
Insert size: 188201; sum-of-contigs
Quality coverage: 4.42x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 61 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs 'N', but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved

1 10439 contig of 10439 bp in length
10540 19520 contig of 8981 bp in length
19621 25661 contig of 6041 bp in length
25762 33211 contig of 7450 bp in length
33312 39008 contig of 5697 bp in length
39109 44927 contig of 5819 bp in length
45028 50794 contig of 5767 bp in length
50895 56215 contig of 5321 bp in length
56316 62553 contig of 6238 bp in length
62654 66956 contig of 4303 bp in length
67057 72631 contig of 5575 bp in length
72732 76352 contig of 3621 bp in length
76453 81462 contig of 5010 bp in length
81563 85655 contig of 4093 bp in length
85756 88387 contig of 2632 bp in length
88488 92170 contig of 3683 bp in length
92271 97116 contig of 4846 bp in length
97217 101536 contig of 4320 bp in length
101637 105091 contig of 3455 bp in length
105192 108459 contig of 3268 bp in length
108560 112286 contig of 3727 bp in length
112387 116086 contig of 3700 bp in length
116187 119641 contig of 3455 bp in length
119742 122405 contig of 2664 bp in length
122506 125633 contig of 3128 bp in length
125734 129128 contig of 3395 bp in length
129229 132328 contig of 3100 bp in length
132429 135207 contig of 2779 bp in length
135308 138331 contig of 3024 bp in length
138432 139316 contig of 885 bp in length
139417 141810 contig of 2394 bp in length
141911 143756 contig of 1846 bp in length
143857 145781 contig of 1925 bp in length
145882 148246 contig of 2365 bp in length
148347 150878 contig of 2532 bp in length
150979 152797 contig of 1819 bp in length
152898 154641 contig of 1744 bp in length
154742 156816 contig of 2075 bp in length
156917 158344 contig of 1428 bp in length
158445 160814 contig of 2370 bp in length
160915 162537 contig of 1623 bp in length
162638 164708 contig of 2071 bp in length
164809 166629 contig of 1821 bp in length
166730 168328 contig of 1599 bp in length
168429 170497 contig of 2069 bp in length
170598 172814 contig of 2217 bp in length
172915 174700 contig of 1786 bp in length
174801 175977 contig of 1177 bp in length

176078 178046 contig of 1969 bp in length
178147 179644 contig of 1498 bp in length
179745 181427 contig of 1683 bp in length
181528 182786 contig of 1259 bp in length
182887 184440 contig of 1554 bp in length
184541 185555 contig of 1015 bp in length
185656 187161 contig of 1506 bp in length
187262 188593 contig of 1332 bp in length
188694 190012 contig of 1319 bp in length
190113 190464 contig of 352 bp in length
190565 191917 contig of 1353 bp in length
192018 193099 contig of 1082 bp in length
193200 194201 contig of 1002 bp in length

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently
* consists of 61 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 10439: contig of 10439 bp in length
10440 10539: gap of 100 bp
10540 19520: contig of 8981 bp in length
19521 19620: gap of 100 bp
19621 25661: contig of 6041 bp in length
25662 25761: gap of 100 bp
25762 33211: contig of 7450 bp in length
33212 33311: gap of 100 bp
33312 39008: contig of 5697 bp in length
39009 39108: gap of 100 bp
39109 44927: contig of 5819 bp in length
44928 45027: gap of 100 bp
45028 50794: contig of 5767 bp in length
50795 50894: gap of 100 bp
50895 56215: contig of 5321 bp in length
56216 56315: gap of 100 bp
56316 62553: contig of 6238 bp in length
62554 62653: gap of 100 bp
62654 66956: contig of 4303 bp in length
66957 67056: gap of 100 bp
67057 72631: contig of 5575 bp in length
72632 72731: gap of 100 bp
72732 76352: contig of 3621 bp in length
76353 76452: gap of 100 bp
76453 81462: contig of 5010 bp in length
81463 81562: gap of 100 bp
81563 85655: contig of 4093 bp in length
85656 85755: gap of 100 bp
85756 88387: contig of 2632 bp in length
88388 88487: gap of 100 bp
88488 92170: contig of 3683 bp in length
92171 92270: gap of 100 bp
92271 97116: contig of 4846 bp in length
97117 97216: gap of 100 bp
97217 101536: contig of 4320 bp in length
101537 101636: gap of 100 bp
101637 105091: contig of 3455 bp in length
105092 105191: gap of 100 bp
105192 108459: contig of 3268 bp in length
108460 108559: gap of 100 bp
108560 112286: contig of 3727 bp in length
112287 112386: gap of 100 bp
112387 116086: contig of 3700 bp in length
116087 116186: gap of 100 bp
116187 119641: contig of 3455 bp in length
119642 119741: gap of 100 bp
119742 122405: contig of 2664 bp in length
122406 122505: gap of 100 bp
122506 125633: contig of 3128 bp in length
125634 125733: gap of 100 bp
125734 129128: contig of 3395 bp in length


```

* 129129 129228: gap of 100 bp
* 129229 132328: contig of 3100 bp in length
* 132329 132428: gap of 100 bp
* 132429 135207: contig of 2779 bp in length
* 135208 135307: gap of 100 bp
* 135308 138331: contig of 3024 bp in length
* 138332 138431: gap of 100 bp
* 138432 139316: contig of 885 bp in length
* 139317 139416: gap of 100 bp
* 139417 141810: contig of 2394 bp in length
* 141811 141910: gap of 100 bp
* 141911 143756: contig of 1846 bp in length
* 143757 143856: gap of 100 bp
* 143857 145781: contig of 1925 bp in length
* 145782 145881: gap of 100 bp
* 145882 148246: contig of 2365 bp in length
* 148247 148346: gap of 100 bp
* 148347 150878: contig of 2532 bp in length
* 150879 150978: gap of 100 bp
* 150979 152797: contig of 1819 bp in length
* 152798 152897: gap of 100 bp
* 152898 154641: contig of 1744 bp in length
* 154642 154741: gap of 100 bp
* 154742 156816: contig of 2075 bp in length
* 156817 156916: gap of 100 bp
* 156917 158344: contig of 1428 bp in length
* 158345 158444: gap of 100 bp
* 158445 160814: contig of 2370 bp in length

```

```

Query Match      100.0%; Score 15; DB 2; Length 194201;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ACCTACGGCATGCAC 15
Db 59787 ACCTACGGCATGCAC 59801

```

```

RESULT 9
LOCUS AP001271
DEFINITION Homo sapiens genomic DNA, chromosome 11q, clone:RP11-826F13,
complete sequence.
ACCESSION AP001271
VERSION AP001271.4 GI:15320468
KEYWORDS HTG.
SOURCE Homo sapiens DNA, clone:RP11-826F13.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
TITLE Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Homo sapiens genomic DNA
REFERENCE Published Only in Database (2000)
AUTHORS 2 (bases 1 to 210402)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
TITLE Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL Direct Submission
Submitted (23-FEB-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT On Aug 27, 2001 this sequence version replaced gi:11138090.
FEATURES
source
1. 210402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="11"
/map="11q"
/clone="RP11-826F13"

```

```

BASE COUNT 50660 a 55101 c 55700 g 48941 t

```

ORIGIN

```

Query Match      100.0%; Score 15; DB 9; Length 210402;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 ACCTACGGCATGCAC 15
Db 210307 ACCTACGGCATGCAC 210321

```

RESULT 10

```

LOCUS AC010674
DEFINITION Homo sapiens chromosome 15 clone RP11-430B1 map 15q21.1, complete
sequence.
ACCESSION AC010674
VERSION AC010674.11 GI:15042780
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 214415)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 214415)
Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
Bloom,S., Dors,M., Dickhoff,R., Harrison,G., James,R., Lasky,S.,
Madan,A., Ratcliffe,A., Shaffer,T. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1999) Multimegabase Sequencing Center, University
of Washington, PO BOX 357730, Seattle, WA 98195, USA
3 (bases 1 to 214415)
Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
Pate,D. and Hood,L.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2001) Multimegabase Sequencing Center, Institute
for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
98105, USA
COMMENT On Jul 31, 2001 this sequence version replaced gi:13878269.
----- Genome Center
Center: Multimegabase Sequencing Center
Center code: UWMSC
Web site: http://chroma.mbt.washington.edu/msg_www
Contact: leerowen@systemsbiology.org
----- Summary Statistics
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 90% of reads
Chemistry: Dye-primer Big Dye; 10% of reads
Assembly program: Phrap; version 0.990399
Note: Data from overlapping BACs AC023906 [drafting center:
UWMSC], AC018902 [drafting center: UWMSC], and AC018903 [drafting
center: UWMSC] were added for finishing
-----
Location/Qualifiers
1. 214415
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15q21.1"
/clone="RP11-430B1"
/clone_lib="RPCI human BAC library 11"
/note="Data from overlapping BACs CTD-2184D3, RP11-327C2,
and RP11-337B11 were added and the consensus sequence was
determined from RP11-430B1 to the extent possible"
1. 11294
/note="overlap with CTD-2184D3 AC023906"
misc_feature

```

```

unsure      65643
             /note="low quality data"
unsure      77163. .77164
             /note="low quality data"
unsure      92425. .92450
             /note="low quality data"
unsure      108954. .108984
             /note="low quality data"
misc_feature 119015. .119515
             /note="sequence data generated from subcloned PCR product
             only"
unsure      126825. .126827
             /note="low quality data"
unsure      146876. .146893
             /note="low quality data"
unsure      165370. .165440
             /note="low quality data"
misc_feature 177221. .214415
             /note="overlap with RP11-327C2"
BASE COUNT  61817 a 47930 c 46801 g 57867 t
ORIGIN
Query Match      100.0%; Score 15; DB 9; Length 214415;
Best Local Similarity 100.0%; Pred. No. 9.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCAC 15
        |||
Db      179173 ACCTACGGCATGCAC 179187

RESULT 11
SME591792      333800 bp      DNA      linear      BCT 05-JUL-2002
LOCUS      Sinorhizobium meliloti 1021 complete chromosome; segment 11/12.
DEFINITION      AL591792 AL591688
ACCESSION      AL591792.1 GI:15075850
VERSION
KEYWORDS
SOURCE
ORGANISM      Sinorhizobium meliloti.
               Sinorhizobium meliloti
               Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
               Rhizobiaceae; Sinorhizobium.
REFERENCE      1 (bases 1 to 333800)
AUTHORS      Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
               Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
               Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
               Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
               Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
               Analysis of the chromosome sequence of the legume symbiont
               Sinorhizobium meliloti strain 1021
               Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE      2 (bases 1 to 333800)
AUTHORS      Gouzy,J.
TITLE      Direct Submission
JOURNAL      Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
               EU Consortium

COMMENT
MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
location/Qualifiers
1. .333800

```

```

/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
complement(233. .1114)
/gene="SMC04043"
complement(233. .1114)
/gene="SMC04043"
/function="miscellaneous: hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC47405.1"
/db_xref="GI:15075851"
/db_xref="SPTREMBL:O92M30"
/translation="MGEAVEWEEFEVLEPASQRIPEFVNSPHSGRHYPOEFLDQSRLLD
PHSIRSEDFHVEDELFRSATFLGAPLIRAHFPRAFLDYKREPELDPRMFDGALPPHA
NISSMRVAGGLGTVPRLVAENMETYRGREFVEQALERIEITYKPYHATLRKLIARTHY
EFGMSLIDCHSMGPNGVHLPGSGRPDEIIGDRYGTSAAELSRVAVELLEOLGYAVA
RNKPYAGGFITEHYGRPTRGHLALQIEINRGLYDEATLIKPGFAALEADLATFTIAA
LARHVEDFGAYLPLAAE"
1305. .1667
/gene="SMC04044"
1305. .1667
/gene="SMC04044"
/function="miscellaneous: not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE 2-COMPONENT RECEIVER DOMAIN PROTEIN"
/protein_id="CAC47406.1"
/db_xref="GI:15075852"
/db_xref="SPTREMBL:Q92M29"
/translation="MTAKILLAEDDNDMRFLVKALEKAGKYKLYSDNGASAYDRLRE
EPFSLLTIDYEMDMDGIELARRATELDPDLKVMFITGFAAVALNPDASKAPKDAKVL
KPEHLRDLVNEVNMKMLAA"
1952. .2026
/gene="trna-VAL_GAC"
1952. .2026
/gene="trna-VAL_GAC"
/product="trna-Val"
/note="codon recognized: GUC; predicted by tRNAscan-SE"
/evidence=not_experimental
2459. .4297
/gene="11VD2 OR SMC04045"
2459. .4297
/gene="11VD2 OR SMC04045"
/EC_number="4.2.1.9"
/function="small molecule metabolism; amino acid
biosynthesis; isoleucine/valine"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE DIHYDROXY-ACID DEHYDRATASE PROTEIN"
/protein_id="CAC47407.1"
/db_xref="GI:15075853"
/db_xref="SPTREMBL:O92M28"
/translation="MPAYRSRTTTHGRNMAGARGLMRATGMKDSDFGKPIIAVNSFT

```

QFVPGVHLKDLGLVAREIEAGVAKENFTIAVDDGIAMGHDMLYSLPSRELIAD
SVEYMNAMGADAMVCISNCDKITPGMMAALRLNIPAVFVSGGPMFAGKVLHGKTH
ALDLVAMVAADDDKVSDEDDVQIERSACPTGSCSGMFTANMCLTEALGLSLPGN
GSTLATADRRKLFVEAGHLIVDLARVYEQEDERVLPRNIATKQAFENAMALDIAMG
GSTNTVLHILAAAYEGEIDFTMDIDRLSRKVPCLSKVAPAKADVHEDVHRAGGIMS
ILGELDKGLINRDCPTVHAETLIDRWDITRTSSDTRKFFRAAPGIPTOVAFS
QEARWDELDTRENGVIRSEVHEPESKDGGLAVLKGNIALDGLCTVKTAGVDESILKFSG
PARVESODAAVKIGILANEIKEMGVVIRYEGPKGPGMOEMLYPTSLKSGLGKAC
ALITDGRFSGTSGLSIGHSVPEANGCTIGLVREGDMIDIDIPNRTISLRVDEAELA
ARTEQDAKGMKPEVQRRKRVTTALAKVAAAFATSADRGAVRDLGDR"
4343. .4400
/note="Sm-5 OR SMC04615
REPEAT SM-5
predicted by Homology"
/evidence=not_experimental
complement(4470. .4874)
/gene="SMC04046"
complement(4470. .4874)
/gene="SMC04046"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC47408.1"
/db_xref="GI:15075854"
/db_xref="SPTREMBL:Q92M27"
/translation="MKTHKSGSCHGRIRREEVDIDLEAGTSRNCSSYCKLRWGWASV
KPEDRLMCEEAGIGDYQFGTMSGHRFCTACGVTPTYGHGYEEIGAFVSINVACLD
GVDPAEFAALPIQYMDGLHNNMWNAPAEFTRHM"
4979. .5024
/note="REP (repetitive extragenic palindromic) element;
predicted by Homology"
/evidence=not_experimental
5027. .5084
/note="Sm-4 OR SMC04584
REPEAT SM-4
predicted by Homology"
/evidence=not_experimental
5094. .5542
/note="Sm-2 OR SMC04873
REPEAT SM-2
predicted by Homology"
/evidence=not_experimental
complement(5586. .6029)
/gene="azu2 OR SMC04047"
complement(5586. .6029)
/gene="azu2 OR SMC04047"
/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE PSEUDODAZURIN (BLUE COPPER PROTEIN)"
/protein_id="CAC47409.1"
/db_xref="GI:15075855"
/db_xref="SPTREMBL:Q92M26"
/translation="MKTMLMLAMCATSAGOANAEYRVEMLNKAADGRVMAFEPAV
IRAQPGDITFVAKDKGNSALMKGAPGAEATWKGINETITVLSKPGVVMYQCAP
HVGGMGIAIVGEPANLEAVKIKYPGKSKAAAEKIFAEIESG"
complement(6076. .6417)
/gene="SMC04048"
complement(6076. .6417)
/gene="SMC04048"

/function="small molecule metabolism; energy transfer;
electron transport"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE CYTOCHROME C PROTEIN"
/protein_id="CAC47410.1"
/db_xref="GI:15075856"
/db_xref="SPTREMBL:Q92M25"
/translation="MTHASSRIAVALPCVTIGLSPRMASAEEDKLAGREIFLERSE

Query Match 100.0%; Score 15; DB 1; Length 333800;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
DB 288772 ACCTACGGCATGCAC 288786

RESULT 12
HSIGVDJ2 787 bp mRNA linear PRI 10-AUG-1995
LOCUS
DEFINITION H.sapiens rearranged IgG VHIII-D-JH-CH3 region.
ACCESSION X81696
VERSION X81696.1 GI:940516
KEYWORDS constant region; immunoglobulin heavy chain; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 787)
AUTHORS Khamlichi,A.A., Aucouturier,P., Preud'homme,J.L. and Coyne,M.
TITLE Structure of abnormal heavy chains in human heavy-chain-deposition
disease
JOURNAL Eur. J. Biochem. 229 (1), 54-60 (1995)
MEDLINE 95262687
PUBMED 7744049
REFERENCE 2 (bases 1 to 787)
AUTHORS Khamlichi,A.A.
TITLE Direct Submission
JOURNAL Submitted (14-SEP-1994) A.A. Khamlichi, CNRS URA 1172,
LAB. IMMUN. AND GENETICS, IBMIG FACULTY OF SCIENCES, F-86022 POITIERS
CEDEX, FRANCE

FEATURES
source location/Qualifiers
1. 787
/organism="Homo sapiens"
/isolate="patient THR"
/db_xref="taxon:9606"
/chromosome="2"
/cell_type="plasma cell"
/tissue_type="bone marrow"
/rearranged
1. .57
/product="immunoglobulin heavy chain"
/note="gamma isotype"
58. .427
/product="immunoglobulin heavy chain V-D-J"
/note="gamma isotype"
428. .750
/product="immunoglobulin heavy chain C3"
/note="gamma isotype"

sig_peptide
V_region
C_region
BASE COUNT 171 a 233 c 225 g 158 t
ORIGIN

Query Match 93.3%; Score 14; DB 9; Length 787;
Best Local Similarity 100.0%; Pred. No. 2e+04;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
Db 149 CCTACGGCATGCAC 162

RESULT 13
AC010298
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTB-163N20, LOW-PASS SEQUENCE
SAMPLING.
AC010298
AC010298.2 GI:6758851
VERSION HTG; HTGS_PHASE0.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 38891)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 38891)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
AUTHORS Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
TITLE On Jan 26, 2000 this sequence version replaced gi:5882635.
JOURNAL
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

* NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 710: contig of 710 bp in length
gap of unknown length
711 1503: contig of 793 bp in length
gap of unknown length
1504 2284: contig of 781 bp in length
gap of unknown length
2285 3002: contig of 718 bp in length
gap of unknown length
3003 3166: contig of 164 bp in length
gap of unknown length
3167 3745: contig of 579 bp in length
gap of unknown length
3746 4319: contig of 574 bp in length
gap of unknown length
4320 5096: contig of 777 bp in length
gap of unknown length
5097 5814: contig of 718 bp in length
gap of unknown length
5815 6228: contig of 414 bp in length
gap of unknown length
6229 6819: contig of 591 bp in length
gap of unknown length
6820 7319: contig of 500 bp in length
gap of unknown length
7320 7678: contig of 359 bp in length
gap of unknown length
7679 8199: contig of 521 bp in length
gap of unknown length
8200 8917: contig of 718 bp in length
gap of unknown length

* 8918 9110: contig of 193 bp in length
gap of unknown length
* 9111 9188: contig of 78 bp in length
gap of unknown length
* 9189 9253: contig of 65 bp in length
gap of unknown length
* 9254 9892: contig of 639 bp in length
gap of unknown length
* 9893 10777: contig of 885 bp in length
gap of unknown length
* 10778 11040: contig of 263 bp in length
gap of unknown length
* 11041 11237: contig of 197 bp in length
gap of unknown length
* 11238 11664: contig of 427 bp in length
gap of unknown length
* 11665 12907: contig of 1243 bp in length
gap of unknown length
* 12908 13764: contig of 857 bp in length
gap of unknown length
* 13765 14507: contig of 743 bp in length
gap of unknown length
* 14508 15239: contig of 732 bp in length
gap of unknown length
* 15240 15636: contig of 397 bp in length
gap of unknown length
* 15637 16504: contig of 868 bp in length
gap of unknown length
* 16505 17834: contig of 1330 bp in length
gap of unknown length
* 17835 17914: contig of 80 bp in length
gap of unknown length
* 17915 18752: contig of 838 bp in length
gap of unknown length
* 18753 18847: contig of 95 bp in length
gap of unknown length
* 18848 19088: contig of 241 bp in length
gap of unknown length
* 19089 19246: contig of 158 bp in length
gap of unknown length
* 19247 19367: contig of 121 bp in length
gap of unknown length
* 19368 19483: contig of 116 bp in length
gap of unknown length
* 19484 19786: contig of 303 bp in length
gap of unknown length
* 19787 20607: contig of 821 bp in length
gap of unknown length
* 20608 20761: contig of 154 bp in length
gap of unknown length
* 20762 20885: contig of 124 bp in length
gap of unknown length
* 20886 20990: contig of 105 bp in length
gap of unknown length
* 20991 21117: contig of 127 bp in length
gap of unknown length
* 21118 22110: contig of 993 bp in length
gap of unknown length
* 22111 23087: contig of 977 bp in length
gap of unknown length
* 23088 24048: contig of 961 bp in length
gap of unknown length
* 24049 24118: contig of 70 bp in length
gap of unknown length
* 24119 24158: contig of 40 bp in length
gap of unknown length
* 24159 24223: contig of 65 bp in length
gap of unknown length
* 24224 24354: contig of 131 bp in length
gap of unknown length
* 24355 24400: contig of 46 bp in length
gap of unknown length
* 24401 24681: contig of 281 bp in length
gap of unknown length

* 24682 25168: contig of 487 bp in length gap of unknown length
* 25169 25245: contig of 77 bp in length gap of unknown length
* 25246 25500: contig of 255 bp in length gap of unknown length
* 25501 26092: contig of 592 bp in length gap of unknown length
* 26093 26876: contig of 784 bp in length gap of unknown length
* 26877 27653: contig of 777 bp in length gap of unknown length
* 27654 27705: contig of 52 bp in length gap of unknown length
* 27706 27772: contig of 67 bp in length gap of unknown length
* 27773 29289: contig of 1517 bp in length gap of unknown length
* 29290 30054: contig of 765 bp in length gap of unknown length
* 30055 31037: contig of 983 bp in length gap of unknown length
* 31038 32220: contig of 1183 bp in length gap of unknown length
* 32221 33342: contig of 1122 bp in length gap of unknown length
* 33343 33411: contig of 69 bp in length gap of unknown length
* 33412 33451: contig of 40 bp in length gap of unknown length
* 33452 33554: contig of 103 bp in length gap of unknown length
* 33555 33607: contig of 53 bp in length gap of unknown length
* 33608 34052: contig of 445 bp in length gap of unknown length
* 34053 34494: contig of 442 bp in length gap of unknown length
* 34495 35154: contig of 660 bp in length gap of unknown length
* 35155 36132: contig of 978 bp in length gap of unknown length
* 36133 38110: contig of 1978 bp in length gap of unknown length
* 38111 38891: contig of 781 bp in length. gap of unknown length.

FEATURES
source 1.38891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTB-163N20"

BASE COUNT 10411 a 9018 c 8505 g 10733 t 224 others
ORIGIN

Query Match 93.3%; Score 14; DB 2; Length 38891;
Best Local Similarity 100.0%; Pred. No. 5.3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCA 14
|||||
Db 34462 ACCTACGGCATGCA 34475

RESULT 14
AF083072 42432 bp DNA linear BCT 06-OCT-1998
DEFINITION Cenarchaeum symbiosum strain B histone H1 DNA binding protein
(hct2), hypothetical protein, lysyl tRNA synthetase (syk), and
hypothetical protein 01 genes, complete cds; 23S ribosomal RNA and
16S ribosomal RNA genes, complete sequence;
glutamate-1-semialdehyde aminotransferase (gsat) and hypothetical
protein 02 genes, complete cds; tRNA-Tyr gene, complete sequence;

triose phosphate isomerase (tim), TATA box binding protein (TF2D),
archaeal family B DNA polymerase B (pol1), dCMP deaminase (dcdt),
ATP-dependent RNA helicase, hypothetical protein 03, menaquinone
biosynthesis protein (mena), and site-specific DNA
methyltransferase genes, complete cds; and unknown genes.
AF083072 GI:3599393
AF083072.1 GI:3599393
Cenarchaeum symbiosum.
Cenarchaeum symbiosum
Archaea; Crenarchaeota; Cenarchaeum.
Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.Y. and
Swanson, R.V.
Genomic analysis reveals chromosomal variation in natural
populations of the uncultured psychrophilic archaeon Cenarchaeum
symbiosum
J. Bacteriol. 180 (19), 5003-5009 (1998)
JOURNAL 98422450
MEDLINE 9748430
PUBMED
REFERENCE 2 (bases 1 to 42432)
AUTHORS Feldman, R.A.
TITLE Direct Submission
JOURNAL Submitted (07-AUG-1998) Genomics, Diversa Corp., 10665 Sorrento
Valley Rd., San Diego, CA 92121, USA
FEATURES
source location/Qualifiers
1.42432
/organism="Cenarchaeum symbiosum"
/strain="B"
/db_xref="taxon:46770"
/clone="60A5"
/note="symbiont of the marine sponge Axinella mexicana"
3.10421
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAC62699.1"
/db_xref="GI:3599394"
/translation="MPAPGEGSLGVAISDGRMYAIGRDLITYRYTMNPBHIA
SAAGAQSFLPGISAPAPGAPGTLSDSDGRHLYVPDENGCVYRFDESPLYRDLGDT
FGSSVYVGSVDVAPRGVYVAPGSLMLVSDADGTHRYELASPYEPAGANRGSFDV
SDMDGSPVGAGFAGLHMYVAGNDTGRVYQYPAGTHQIOEAAGPRLSAVLKDGTL
RAAFDGTVDAGSVQPGMTLRDGHGNTGIPLLAGAADSDVMTFVVEKDRAEAAA
YGDQSLHVPAAALAGTGGGPPFVDFSGSLASLYRHERPQGEEMAREKSDRYALT
VTAGGSQMHVGGAGGNTWYDLGTPHDTTGVVAGSDILPAYPSAGRNVPSTIGIAE
SDGMRLEAANRGDRIPMYQLDSRPYDIGSASLEGLFTGFQSGIAFSDGTRMFAALL
TENAIRQYDLGEPYDIRGAGNAGYDLDPLHPLGLFLTSGVHFSPTGRMFVGGI
SDAFEDANANRDVNVNLRHREDLSTPEVDLTAEKYDTYESTGPAGDLDELSTSPDGR
LYTLSSERVSSSEYTTIRAOYWLPEPYDVTPIYHVPSEFNASOGNLADEGMAFSFDG
TRLVLTGHGQTNAKLHNLNPPFDVGTAVFHDHGRFRPGPASETEASGTSLSADGRM
FLSDRGKGAISQYTLVAPFVDFEAFSDVSADGQLDVGAQDALPGGLAFSPGTRLMVG
GMDRSVHMYSLNTPFDLGAFAHAASFGVGDVSDPLIEGETDRVYEHRCIYIPVSAL
YDLGAPYDISGPAYSGIFDAGSIRDAVAGGSMFLIEGETDRVYEHRCIYIPVSAL
DGPALVSAADARVGAELVDFRAVDVGGIDPGGVRIVDAAGPLPGVVISDAVPIGED
PGVARESLSDAEVLAVSGYAEPSLVFGRHAVPGAGGTEPQIGNATELVGSIPTL
DEGTLTGAAFSADGTYVFLSDGPTGRVYPSLNPDISSAAPGCFVYVPGVSDIA
FSADGRNMLVADETGGIHYRLARSPYEIGTDFIKSSLGEVETFSAPARODLAGIAF
SHDGMIMLAAGSGSVHRYSLPSYAVSGAKYEETAMIGSPSGLEFSSDGLRMFVPD
AGSETAAVYGLAAPYGIAGEAPLPLFLGVGAEEATLSPDGRHILVPGRGLSQYSLF
STNLELCAEPRGIDGSGCEDGIYAFESPGREGVSLAASITAADPGICGELHGFAGP
MPAPVMEQVTLDSREGTLRVLRDVTVDVTRPKMWEVDSDGQTLNANSTLANNAEN
SNILFLRDDAAGAKISGYTSPVFRWSSPFLGTGATRPHTLGFGLVRLADIYDASG
DVPSPSGIEFSDDGMRMEVGTGICPGINIFTLAPEDITLPHKSGSTNIGLSVSDLA
FANNNGSLVFLVDVGLRVYALDDYNNVTGQKERRITLDTTGIPNSIYTSPDGLS
OEVAYDDRIDLYLVGSPNDISSTTEIIPYSLPRDPPTGMDFTPDGRMFLSTENGID
QYLLSEPFVAVTTSVFLRTIPIDGAGEIRFVNDGRGLFVGADGIQRHMLIYPYGAS
TSLETVRDGVTGDPGPNPAGAEIRLACTFNASNVQSPSGIEFSGDGTGMEFVTGFG
AAGVNEFSLAPEDTTLVPEVHELHIDIGQAPVDAFAFEDGRTLLLAADGTLDFEYSLA
GDAYDIGEASRTPOVPFEDAAGAVGAFYOPRDSIIAIFDGRIDQYVYVPEFVS
PLTRPGTPTGIDFAPDGRWMLSTENGIDQYLLSIPFDVRSITVGTIPDVGEGMOF
ADNGRALFLADSEGLIYNDLEDYALDGNITISVEFSFGSVMYVLEYDTRKRVSYEL
EFPEDVSSRTRADTLIDPQIDSPRHVAVSMGMNHIYITNSVFGEDDTIHSYGISNNDI
SSASYIGEGLPEPVINGIDFSNNGRRMFLIGNGFDYQYIHDYMLGTRYDISSKSL

CDS

DTYAIPGVYEPAGLDSEFDRLSMFTISTAGSVYRYGLDDPIVETMDYQESFRLPV
SAADNSIDLAFGSSGLNAVISHEGDITLYSFVLDPYGAELDIDRLLEPLGVPTGF
EFSDNGRQLYIGAFRDSQSSPTLPAQLQRYELGIPYDLASAVEAQLGIEDFPFNG
MRANGSLAGLHVPPDGSILFRAGNAERTVISYDMDSHDLDITLSEFSKPDVGOSTPN
IRDMDISPDGMFLYLQGDVLDWMYNTDSYSLDAPAYAGTLDLEPEDVTPRGISFSD
GTSLFMTGEDVDHIHEYALNEPWDIRNALIAGSLISAVNGAPRGLDISEDGTTAHTM
RGRDFDTGPASLVNHILPGQYSTLTDPAFAYPVEEGAPGLAFSDDGMRMFVAGVN
NHLRQYNLSPDYDTENAEHFISTDILTADRGPPTGLVFSDENDFFSTGARAQFVRQFTT
NRPYDASTITLSDNGLYKVSVDLPSCGTRFPTDGMKFIISGQETAMITYOYSLPSPYDT
SGAVRDRVEIVAGLFRNAGLSVGLNPSPPSGDFSEDMELTYTGSLVHRYFLPSPY
GLEDAAYGSGFHTRESTPLGVVVRGDA MFVAGDSTDSILKYSLNAQPVGNITHADTR
AGIADRAEIVGAMADTRAELDGDADVHKSVIDVFPISEGITVGRALYPEDAAILD
DGANATHNRVYIIVHDITEGDAPSIHDEPIAVGIYALGPMDTIAYVDLHRLAVSASLS
GDSPSASDASGVVAESRRNAVDRPGVEERIGHVSLAADRPAVDMMDTDSAGVYD
RSPDDGPAYSDRSALGLARMAADRPVDDMDTDSAGVYDRSPDGPASIDRSALGLA
RMAADRPVDDMDTGSAGVYDRSPDGPASIDRSALGLARMAADRPVDDMDTGS
STSRLGVPDRPEIVERHSLAASVYLSGGDSPSVADGHVSEGRDGDGDRPIDERIV
IKISYSRGADAPRVEDAMETSGVTAYSRGADAPRVEDAMETSGVTVPRRSTMADPT
VADHSLARTASISEGDSPTFAEARADTVGDIDEVDAPTVAADHSLARAASISEGDS
PTFAEVRADTVGDIDEVDAPAVAEERLAVLGQAPDSGVWMDTVGLDISEISGDPVP
EPRVVRPGGGGGGSSNRGLPEHGGGYIDEFRIDRLVLFNGTIDVLAESGKDLI
RPVFRPEGSENFIDMEVLFTARGPEISTAYYNRAGILMGIDCGELIMTDTTYS
IFGDEIYHVERIDAFNGMVISLDPGLDGVSVSLRDNHGIPLAQHRHAKYEILLDAA
ENRPLSVSTDPKVEDPSPVQHTESLQMDPEPVESEPLMDSEPEVLEPVQHLES
MDPEPEVLEPVQHLEPVQSGSPVPVQGPSPVESGIAATLMQFLSGLDALGLADPDVG
SVQKTS"

gene 10625. .11434
/gene="hc2"
CDS 10625. .11434

/gene="hc2"
/note="similar to eukaryotic histone H1 DNA binding protein"
/codon_start=1
/transl_table=11
/product="histone H1 DNA binding protein"
/protein_id="AAC62700.1"
/db_xref="GI:3599395"
/translation="MHGIEGGRGDMSENFVAFCAARGVTKDEMXYDGRVHKECH
ARHGQIRFPNPEVEQRYAELKVDLIQMRNOLAEMNRASGDGVHSSATSAAEAQHR
AELKVLQVOMRNOLEMNRRKAPGKPARKKAAGTKARRKSGKTVRRKTGTAGKAG
ARRKTVKRTARRKTTAKKAAGRKAGARRKATYKRTVHKKIGVRRKTTARRTAGKSTV
RRKSTVKRTVHRKTGKKAVVRKSTVKRTARRPAGRKTPGRAARRAGAKRR"
11478. .13046
CDS /codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC62701.1"
/db_xref="GI:3599396"
/translation="MOSLGRLDEACAEISRSLEYESPTAGDVRTETIRACTKYSLR
IPKNREILATARGODFDRRLRPLLKKPVKTASGAVIAVMPMPYACPHGRCYCPGGE
ASNTPNSYTGGEPIAAGAMNSGYDPEEQVRAGLRLRAHGDVAKLEIVVGGTFLFM
POEYQEWFKSCYDALNGSASAGMEBAKHNETAVHRNVGLTIETKPDYCRTEHVDA
IGFGATRVEIGVOSLREEVYLVRNGHYODVTFESFAARDYAKVAAHMMGLPGAT
PEGDIEDLRMLFEDPALRPDMLKVYPALVVRGTPMEEYSRGEYSPTYEEVIRVLSE
AKARVPRMARIMRVOREIHPDEIVAGPRSGNLROLVHKRLQEOGRRCRCIRCEAGLA
GRTVPQKLRIDRADYSASGGRESFISLVDDDAIYGFVRLRKPSSGAHREPVTPESC
IRELHVYGRSLGLGERGCIQHSGLGRLVSEASARELGAQLLVISA VGTGRGYRR
LGYSRTGPYMGKVL"

gene 13046. .14620
/gene="syk"
CDS 13046. .14620

/gene="syk"
/codon_start=1
/transl_table=11
/product="lysyl tRNA synthetase"
/protein_id="AAC62702.1"
/db_xref="GI:3599397"
/translation="METIGRGWIDKLAHELVEREALGRDTEMINVESGLGASGIPH
MGLGDVAVRAYGVLAVGDMGHSFRLIAYFDDLGLRKVPEGMPSSLIEHIIARPYSAI
PDYRGCHDSYGMHSGLLLEGDALGIEYDFRRARDTVRGDLAEQIHRILSNSSVIG
EKIAEMVGOEKFRSSLPYFAVCEQCGMYTAESVEYLADSRKVRYRCGDAEYVGKRIA
GCGHEGEADTGGAGGKLAWKVEFAARWQAFDVRFEAYGKDIMDSVRINDWVSEILSS
PHPHHTRYEMFLDKGKKISKSSGNVVTPOKWLRYGTFPOSILLMYKRITGARELGLE

CDS

rRNA
rRNA
gene
CDS

DVPSLMDEYDGLQREYFAGGGRGKAREAKNRGLFEYTNLLLEAQEGPRHAGYRLVE
LSRLFERNTERVTKKLVEYGVIDGSPGIERLIALAGNYADDMYSAEREVELDGAT
RGALSELAEMLGSAPEGGIQDVIYGVAKSHGVPRDFFKALYRIILLDASSGRIGPFI
EDIGREKVAGMIRGRL"
complement(17329. .18213)
/codon_start=1
/transl_table=11
/product="hypothetical protein 01"
/protein_id="AAC62703.1"
/db_xref="GI:3599398"

/translation="MESAGEQAPGVVLHDYLSKLOQYSGRDTILYATNMWTDPEHTPN
EALITNGDLYGFMRMMDLTKKLDLILHSPGSAESAESIYTYLHAKYDRIYIIPY
AAMSASMLACASNSLYMCKHSSIGPADPQIFPTKIGMOIMSAQLIDELQEVQVS
EKHPGRLGAWPLLGQYPPGLVOKCISQKLAELVQKWLDDHMFAGESDAEKS
SGMLASPGKYXSHGRYISRECRGIGLKITDLEADQEFQDLTILSVSHAADILSQFTPI
NKIITANHLGNSVISKPSI"

18645. .21639
/product="23S ribosomal RNA"
21771. .23243
/product="16S ribosomal RNA"
23558. .24862
/gene="gsat"
23558. .24862
/gene="gsat"
/function="heme biosynthesis"
/codon_start=1
/transl_table=11
/product="glutamate-1-semialdehyde aminotransferase"
/protein_id="AAC62704.1"
/db_xref="GI:3599399"

/translation="MDLEREYRAKTRGSAGIFARSRRYHGVSHNIRIYEPFVTR
SARGKHLVDVDGNKTYTDYMMGHWSLILGHAPQVRSVAVGQLRRGWIHGTAPEPTRL
SEIRGAVKAAEKIRYVTSGETEAVMYAARMARARTGKYIAKVIVGWHGYASGLIKSV
NMPYDVPESGGLVDEEHTVSIPYNNLEGSLEALRRAGGDLACVIVEPMLGGCGCIPAE
PDYLRGIOEFVHSGKALFILDEIVTGFREFDGCAYKKMGUDPVVALGKIVGGGEP
VVCCKDEVMCISDTGAHARTERAYIGGTFSANPATMTAGAAALGALRERRGTLYPR
NSMGDARARLSRIFFDGRVAVTGRGSLFMTHTTPDGARRISSAADAACDVHLLHRYH
LDMITRDGIFFLPGKLGAISAHSRADLGAMYASASERFAGL"
24913. .25728

CDS

Query Match 93.38; Score 14; DB 1; Length 42432;
Best Local Similarity 100.0%; Pred. No. 5,le+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCTACGGCATGCAC 15
Db 13377 CCTACGGCATGCAC 13390

RESULT 15
AL357123 72354 bp DNA linear PRI 09-MAR-2001
LOCUS
DEFINITION Human DNA sequence from clone RP11-287E6 on chromosome Xq26.3-27.3,
complete sequence.
ACCESSION AL357123
VERSION AL357123.11 GI:13274730
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Wray, P.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,

COMMENT

CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 12, 2001 this sequence version replaced gi:13234914.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP, Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
RP11-287E6 is from the library RPCL-11.1 constructed by the group of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-287E6 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP6-27P15 is at 72255 in this sequence. The true right end of clone RP11-65G20 is at 100 in this sequence.

FEATURES

source
1..72354
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q26.3-27.3"
/clone="RP11-287E6"
/clone_lib="RPCL-11.1"
/complement(1..100)
/note="match: STS: Em:G60993
match: GSS: Em:AQ200790"
complement(916..1482)
/note="match: STS: Em:HSC19C12"
2401..2612
/note="L2 repeat: matches 2482..2669 of consensus"
2762..2804
/note="MER3 repeat: matches 7..49 of consensus"
2783..3166
/note="match: GSS: Em:AQ033161"
2909..3008
/note="L1ME1 repeat: matches 5840..5947 of consensus"
4413..4608
/note="MLT1H repeat: matches 345..545 of consensus"
4774..4900
/note="MLT1G repeat: matches 43..177 of consensus"
5067..5129
/note="L2 repeat: matches 1919..1981 of consensus"
7760..7958
/note="L2 repeat: matches 2565..2750 of consensus"
7996..8111
/note="L2 repeat: matches 2575..2698 of consensus"
9225..9386
/note="L1MC5 repeat: matches 7774..7931 of consensus"
9537..10038
/note="L1ME1 repeat: matches 5431..5954 of consensus"
10762..10832
/note="L2 repeat: matches 1728..1798 of consensus"
11832..12035
/note="MIR repeat: matches 34..250 of consensus"

repeat_region 12703..13123
/note="HSMAR1 repeat: matches 1..439 of consensus"
repeat_region 13156..13426
/note="L1M4 repeat: matches 3062..3342 of consensus"
repeat_region 13465..14328
/note="L1M4 repeat: matches 2128..2988 of consensus"
repeat_region 14546..15060
/note="L1MC4 repeat: matches 7285..7835 of consensus"
repeat_region 15550..15677
/note="HAL1 repeat: matches 173..312 of consensus"
repeat_region 15763..16499
/note="L1PA10 repeat: matches 5404..6153 of consensus"
repeat_region 16490..16843
/note="HAL1 repeat: matches 408..759 of consensus"
repeat_region 16966..17089
/note="62 copies 2 mer tt 71% conserved"
repeat_region 17051..17102
/note="13 copies 4 mer ctcc 92% conserved"
repeat_region 17115..17414
/note="AluSx repeat: matches 1..301 of consensus"
repeat_region 17578..18034
/note="L1PA13 repeat: matches 5695..6152 of consensus"
repeat_region 18379..18691
/note="AluSx repeat: matches 1..310 of consensus"
repeat_region 20868..21162
/note="AluSg repeat: matches 1..296 of consensus"
repeat_region 21417..21600
/note="MIR repeat: matches 35..212 of consensus"
repeat_region 22265..22415
/note="MIR repeat: matches 2..143 of consensus"
repeat_region 22581..22664
/note="21 copies 4 mer atag 78% conserved"
repeat_region 24520..24721
/note="MER5B repeat: matches 6..178 of consensus"
repeat_region 25456..27803
/note="L1 repeat: matches 2387..4703 of consensus"
repeat_region 28250..28835
/note="L1 repeat: matches 4117..4755 of consensus"
repeat_region 29493..29573
/note="L1 repeat: matches 4023..4106 of consensus"
repeat_region 29574..29870
/note="AluSg repeat: matches 1..297 of consensus"
repeat_region 29871..29964
/note="L1 repeat: matches 3939..4023 of consensus"
repeat_region 30234..30588
/note="THE1A repeat: matches 1..354 of consensus"
repeat_region 31405..31491
/note="MLT1I repeat: matches 1..85 of consensus"
repeat_region 32756..33050
/note="AluSg repeat: matches 1..295 of consensus"
repeat_region 33150..33283
/note="L2 repeat: matches 808..951 of consensus"
repeat_region 33470..33821
/note="MLT1F repeat: matches 201..539 of consensus"
repeat_region 33881..34193
/note="AluSx repeat: matches 1..312 of consensus"
repeat_region 34199..34368
/note="MLT1F repeat: matches 10..172 of consensus"
repeat_region 34436..34822
/note="MLT1A1 repeat: matches 1..365 of consensus"
repeat_region 34824..34960
/note="L1M3 repeat: matches 5464..5597 of consensus"
repeat_region 34958..35119
/note="L1M4 repeat: matches 3673..3836 of consensus"
repeat_region 35159..35248
/note="L2 repeat: matches 1249..1337 of consensus"
repeat_region 35249..35553
/note="AluSx repeat: matches 1..305 of consensus"
repeat_region 35554..35948
/note="L2 repeat: matches 1337..1739 of consensus"
repeat_region 35957..36082
/note="L2 repeat: matches 1945..2077 of consensus"
repeat_region 36358..37198

```

repeat_region /note="L1MB6 repeat: matches 5318. .6174 of consensus"
37209. .37447
repeat_region /note="L1PA15 repeat: matches 5918. .6157 of consensus"
37475. .37531
repeat_region /note="L1P repeat: matches 2395. .2451 of consensus"
37571. .39494
repeat_region /note="L1PA13 repeat: matches 523. .2465 of consensus"
39502. .40031
repeat_region /note="L1M4 repeat: matches 4793. .5331 of consensus"
40379. .41594
repeat_region /note="L1M4 repeat: matches 2614. .3869 of consensus"
41599. .41684
repeat_region /note="L1M4 repeat: matches 1225. .1319 of consensus"
41685. .41978
repeat_region /note="AluSg repeat: matches 1. .306 of consensus"
41979. .41992
repeat_region /note="L1M4 repeat: matches 1213. .1225 of consensus"
41998. .42194
repeat_region /note="AluSp repeat: matches 116. .313 of consensus"
42199. .43409
repeat_region /note="L1M4 repeat: matches -12. .1191 of consensus"
43430. .43525
repeat_region /note="L2 repeat: matches 2399. .2495 of consensus"
43606. .43736
repeat_region /note="L2 repeat: matches 2614. .2750 of consensus"
44810. .46062
repeat_region /note="L2 repeat: matches 1467. .2749 of consensus"
46166. .46419
repeat_region /note="MIR repeat: matches 2. .255 of consensus"
46662. .46809
repeat_region /note="L1M4 repeat: matches 4611. .4760 of consensus"
46910. .48311
repeat_region /note="L1PA10 repeat: matches 4755. .6158 of consensus"
48340. .48375
repeat_region /note="L1ME2 repeat: matches 5386. .5419 of consensus"
48376. .48668
repeat_region /note="AluSx repeat: matches 1. .295 of consensus"
48669. .49381
repeat_region /note="L1ME2 repeat: matches 5419. .6164 of consensus"
49857. .49964

```

Query Match 93.3%; Score 14; DB 9; Length 72354;
 Best Local Similarity 100.0%; Pred. No. 4.3e+03;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGCGCATGCAC 15
 |||||||||||||
 Db 11377 CCTACGCGCATGCAC 11390

Search completed: June 26, 2003, 03:47:32
 Job time : 428.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-25

Perfect score: 15

Sequence: 1 acctacgcgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	20	AA32399	Ab3 variable heavy
2	15	100.0	115	24	ABK71361	Thrombopoietin (TP
3	15	100.0	115	24	ABA04558	Oligonucleotide #2
4	15	100.0	351	24	ABK71358	DNA encoding throm
5	15	100.0	351	24	ABA04555	Human MPL #1 codin
6	15	100.0	432	24	ABK71369	DNA encoding throm
7	15	100.0	432	24	ABA04563	Human coding seque
8	15	100.0	717	20	AA317988	Anti-Mpl scFv 12B5
9	15	100.0	823	24	ABK71387	DNA encoding throm

C	10	15	100.0	2908	24	ABL56609	Nucleotide sequenc
	11	14	93.3	1575	21	AAA55191	Cenarchaeum symbio
	12	14	93.3	42432	21	AAA55187	Cenarchaeum symbio
	13	13.4	89.3	65	24	ABN57826	Mouse spliced tran
C	14	13.4	89.3	292	24	ABL71855	Corn tassell-derive
	15	13.4	89.3	372	22	AAH68618	Human anti-Rh(D) c
	16	13.4	89.3	375	22	AAH68623	Human anti-Rh(D) c
	17	13.4	89.3	429	21	AAA13930	Human anti-Rh(D) c
	18	13.4	89.3	429	21	AAA13932	Human PTHrP monocl
	19	13.4	89.3	429	21	AAA13933	Human PTHrP monocl
	20	13.4	89.3	429	21	AAA13935	Human PTHrP monocl
	21	13.4	89.3	429	21	AAA13937	Human PTHrP monocl
C	22	13.4	89.3	471	20	AA391539	Human PTHrP monocl
	23	13.4	89.3	714	24	ABK36221	Human PTHrP monocl
	24	13.4	89.3	936	22	AAH43565	Human PTHrP monocl
	25	13.4	89.3	1157	21	AAA39072	Human PTHrP monocl
	26	13.4	89.3	1320	20	AA39072	Human PTHrP monocl
	27	13.4	89.3	1329	18	AA39072	Human PTHrP monocl
	28	13.4	89.3	1379	22	AAH43561	Human PTHrP monocl
	29	13.4	89.3	1612	23	AA39072	Human PTHrP monocl
C	30	13.4	89.3	1689	22	AA39072	Human PTHrP monocl
	31	13.4	89.3	1743	22	AA39072	Human PTHrP monocl
C	32	13.4	89.3	2253	24	ABK84330	Human PTHrP monocl
C	33	13.4	89.3	2253	24	ABK84330	Human PTHrP monocl
C	34	13.4	89.3	2253	24	ABK84330	Human PTHrP monocl
	35	13.4	89.3	2472	22	ABA89120	Human PTHrP monocl
C	36	13.4	89.3	2939	21	AA39072	Human PTHrP monocl
C	37	13.4	89.3	2939	21	AA39072	Human PTHrP monocl
C	38	13.4	89.3	3054	23	ABL04210	Human PTHrP monocl
	39	13.4	89.3	3735	19	AA39072	Human PTHrP monocl
	40	13.4	89.3	3735	21	AA39072	Human PTHrP monocl
	41	13.4	89.3	3735	24	AA39072	Human PTHrP monocl
	42	13.4	89.3	3735	24	AA39072	Human PTHrP monocl
C	43	13.4	89.3	3735	24	AA39072	Human PTHrP monocl
	44	13.4	89.3	4916	23	ABL18373	Human PTHrP monocl
	45	13.4	89.3	6594	23	ABL15639	Human PTHrP monocl

ALIGNMENTS

RESULT 1	
AA32399	
ID	AA32399 standard; DNA; 15 BP.
AC	AA32399;
XX	
DT	17-JUN-1999 (first entry)
XX	
DE	Ab3 variable heavy (VH) chain CDR1 encoding DNA.
XX	
KW	Agonist antibody; thrombopoietin receptor; TP0-R; thrombopoietin; DIC;
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;
KW	myelodysplasia; myelotoxic chemotherapy; leukaemia; tumour; MUSK; CDR;
KW	neuromuscular; muscular dystrophy; complementarity determining region;
XX	variable heavy chain; variable light chain; VH; VL; SS.
OS	Homo sapiens.
XX	
PN	WO9910494-A2.
PD	04-MAR-1999.
XX	
PF	21-AUG-1998; 98WO-US17364.
XX	
PR	25-AUG-1997; 97US-0918148.
XX	
PA	(GETH) GENENTECH INC.
XX	
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;
XX	
DR	WPI; 1999-204666/17.

DR P-PSDB; AAY06699.
XX
PT New thrombopoietin receptor agonist antibodies - useful for
PT treating immunological or hematological disorders
XX
PS Claim 10; Page 78; 86pp; English.
XX
CC The invention relates to an agonist antibody (Ab) which binds to a
CC thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
CC be used in the same way and for the same indications as thrombopoietin
CC (TPO). They can stimulate proliferation, differentiation or growth of
CC megakaryocytes. They may also be able to stimulate megakaryocytes to
CC increase platelet production. They can be used for treating
CC immunological or hematopoietic disorders, especially thrombocytopenia.
CC Thrombocytopenia -associated bone marrow hypoplasia (e.g. aplastic anemia
CC following chemotherapy or bone marrow transplant) may be effectively
CC treated with the antibody compounds as well as disorders such as
CC disseminated intravascular coagulation (DIC), immune thrombocytopenia
CC (HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
CC congenital thrombocytopenia, thrombotic thrombocytopenia and
CC myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
CC treatment of solid tumours or leukaemia, myeloablative chemotherapy for
CC autologous or allogeneic bone marrow transplant, myelodysplasia,
CC idiopathic aplastic anemia, congenital thrombocytopenia, and immune
CC thrombocytopenia. The antibodies which bind to the MUSK receptor can be
CC used for improving neuromuscular function in a patient, e.g. in muscular
CC dystrophy. The products can also be used for detection and diagnosis. The
CC antibodies have a longer half-life than the natural ligand for the TPO-R.
CC Sequences AAX32387-X32413 represent DNA fragments encoding the CDR1,
CC CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
CC (VL) chains of antibodies Ab1 to Ab6.
XX
SQ Sequence 15 BP; 4 A; 6 C; 3 G; 2 T; 0 other;

Query Match 100.0%; Score 15; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
 |||||
DB 1 ACCTACGGCATGCAC 15

RESULT 2
ABK71361/c
ID ABK71361 standard; DNA; 115 BP.
XX
AC ABK71361;
XX
DT 30-JUL-2002 (first entry)
XX
DE Thrombopoietin (TPO) agonist antibody associated polynucleotide #2.
XX
KW Modified antibody; thrombopoietin; TPO; agonist;
KW TPO receptor; platelet reduction-associated blood disease;
KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
KW ds.
XX
OS Synthetic.
XX
PN WO200233072-A1.
XX
PD 25-APR-2002.
XX
PF 22-OCT-2001; 2001WO-JP09259.
XX
PR 20-OCT-2000; 2000JP-0321821.
PR 17-APR-2001; 2001WO-JP03288.
PR 12-SEP-2001; 2001JP-0277314.
XX
PA (CHUS) CHUGAI SEIYAKU KK.
XX
PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

XX
DR WPI; 2002-383513/41.
XX
PT Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
PT leukaemia
XX
PS Example 7; Page 178; 213pp; Japanese.
XX
CC The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives
CC and/or remedies for platelet reduction associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukaemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence represents thrombopoietin (TPO) agonist antibody associated
CC polynucleotide used in the creation of the modified antibody described in
CC the invention.
XX
SQ Sequence 115 BP; 25 A; 39 C; 33 G; 18 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 115;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
 |||||
DB 68 ACCTACGGCATGCAC 54

RESULT 3
ABA04558/c
ID ABA04558 standard; DNA; 115 BP.
XX
AC ABA04558;
XX
DT 15-FEB-2002 (first entry)
XX
DE Oligonucleotide #2.
XX
KW Cytostatic; antiinflammatory; antianaemic; vasotropic;
KW antibody; signal transmissio; cancer; inflammation; hormonal disorder;
KW leukaemia; lymphoma; aplastic anaemia; skeletal malformation; ss.
XX
OS Synthetic.
XX
PN WO200179494-A1.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-JP03288.
XX
PR 17-APR-2000; 2000JP-0115246.
PR 20-OCT-2000; 2000JP-0321821.
PR 20-OCT-2000; 2000JP-0321821.
PR 12-MAR-2001; 2001WO-JP01912.
XX
PA (CHUS) CHUGAI SEIYAKU KK.
XX
PI Fukushima N, Tsuchiya M, Oheda M, Uno S, Kikuchi Y, Ohtomo T;
XX
DR WPI; 2002-066368/09.
XX
PT Antibodies for treatment of diseases associated with cell
PT proliferation, hormonal disorders and cytokines comprise agonist
PT activity to signal transmissio across cell membranes -
XX
PS Example 7; Page 157; 173pp; Japanese.
XX
CC The present invention relates to modified antibodies. The antibodies

CC contain two or more H chain V domains and two or more L chain V domains
CC of a monoclonal antibody (Mab) which is capable of transmitting a signal
CC across the cell membrane by cross-linking a cell surface molecule, where
CC the antibodies can serve as signal transmission agonists. The antibodies
CC are useful for treatment and prevention of a broad range of disorders in
CC which signal transmission is implicated, such as cancer, inflammation,
CC hormonal disorders, leukaemia, malignant lymphoma, aplastic anaemia, and
CC skeletal malformations. The present sequence was used in an example from
CC the present invention.
XX
SQ Sequence 115 BP; 25 A; 39 C; 33 G; 18 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 115;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
Db 68 ACCTACGGCATGCAC 54

RESULT 4
ABK71358
ID ABK71358 standard; DNA; 351 BP.
XX
AC ABK71358;
XX
DT 30-JUL-2002 (first entry)
XX
DE DNA encoding thrombopoietin agonist antibody associated protein #14.
XX
XX Modified antibody; thrombopoietin; TPO; agonist;
KW TPO receptor; platelet reduction-associated blood disease;
KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
KW gene; ds.
XX
OS Homo sapiens.
XX
XX WO200233072-A1.
XX
PN 25-APR-2002.
XX
PD 22-OCT-2001; 2001WO-JP09259.
XX
PF 20-OCT-2000; 2000JP-0321821.
XX
PR 17-APR-2001; 2001WO-JP03288.
PR 12-SEP-2001; 2001JP-0277314.
XX
XX (CHUS) CHUGAI SEIYAKU KK.
XX
PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;
XX
XX WPI; 2002-383513/41.
DR P-PSDB; ABG35320.
XX
XX Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
PT leukaemia -
XX
PS Example 7; Page 175-176; 213pp; Japanese.
XX
CC The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives
CC and/or remedies for platelet reduction-associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukaemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence encodes a thrombopoietin (TPO) agonist antibody associated
CC protein.
XX

SQ Sequence 351 BP; 75 A; 90 C; 113 G; 73 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 351;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
Db 91 ACCTACGGCATGCAC 105

RESULT 5
ABA04555
ID ABA04555 standard; DNA; 351 BP.
XX
AC ABA04555;
XX
DT 15-FEB-2002 (first entry)
XX
DE Human MPL #1 coding sequence.
XX
XX Human; cytostatic; antiinflammatory; antianaemic; vasotropic;
KW antibody; signal transmission; cancer; inflammation; hormonal disorder;
KW leukaemia; lymphoma; aplastic anaemia; skeletal malformation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 1..351
FT /*tag= a
FT /partial
FT /product= "Human MPL #1"
FT /note= "No start or stop codon given"
XX
XX WO200179494-A1.
XX
PD 25-OCT-2001.
XX
PF 17-APR-2001; 2001WO-JP03288.
XX
PR 17-APR-2000; 2000JP-0115246.
PR 20-OCT-2000; 2000JP-0321821.
PR 20-OCT-2000; 2000JP-0321822.
PR 12-MAR-2001; 2001WO-JP01912.
XX
XX (CHUS) CHUGAI SEIYAKU KK.
XX
XX Fukushima N, Tsuchiya M, Oheda M, Uno S, Kikuchi Y, Ohtomo T;
XX
XX WPI; 2002-066368/09.
DR P-PSDB; AAM47640.
XX
XX Antibodies for treatment of diseases associated with cell
PT proliferation, hormonal disorders and cytokines comprise agonist
PT activity to signal transmission across cell membranes -
XX
PS Example 7; Page 155-156; 173pp; Japanese.
XX
XX The present invention relates to modified antibodies. The antibodies
CC contain two or more H chain V domains and two or more L chain V domains
CC of a monoclonal antibody (Mab) which is capable of transmitting a signal
CC across the cell membrane by cross-linking a cell surface molecule, where
CC the antibodies can serve as signal transmission agonists. The antibodies
CC are useful for treatment and prevention of a broad range of disorders in
CC which signal transmission is implicated, such as cancer, inflammation,
CC hormonal disorders, leukaemia, malignant lymphoma, aplastic anaemia, and
CC skeletal malformations. The present sequence was used to illustrate the
CC present invention.
XX
SQ Sequence 351 BP; 75 A; 90 C; 113 G; 73 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 351;
Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
 |||||
 DB 91 ACCTACGGCATGCAC 105

RESULT 6
 ABK71369
 ID ABK71369 standard; DNA; 432 BP.

XX AC ABK71369;

DT 30-JUL-2002 (first entry)

DE DNA encoding thrombopoietin agonist antibody associated protein #17.

XX Modified antibody; thrombopoietin; TPO; agonist;
 KW TPO receptor; platelet reduction-associated blood disease;
 KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
 KW gene; ds.

XX OS Homo sapiens.

PN WO200233072-A1.

PD 25-APR-2002.

PF 22-OCT-2001; 2001WO-JP09259.

PR 20-OCT-2000; 2000JP-0321821.

PR 17-APR-2001; 2001WO-JP03288.

PR 12-SEP-2001; 2001JP-0277314.

PA (CHUS) CHUGAI SEIYAKU KK.

PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

DR WPI; 2002-383513/41.

DR P-PSDB; ABG35323.

PT Degraded thrombopoietin agonist antibodies containing H and L chain V

PT domains of monoclonal antibody, useful in preventives and/or remedies

PT for blood diseases, thrombocytopenia following cancer chemotherapy or

PT leukaemia

PS Example 7; Page 181-182; 213pp; Japanese.

XX The invention describes a modified antibody comprising at least 2 heavy

CC chain variable domains and 2 or more light chain variable domains of an

CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing

CC the TPO receptor to crosslink. The antibodies are useful in preventives

CC and/or remedies for platelet reduction-associated blood diseases.

CC thrombocytopenia following cancer chemotherapy or leukaemia. The

CC antibody can act as a TPO signal transduction agonist by transducing a

CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.

CC This sequence encodes a thrombopoietin (TPO) agonist antibody associated

CC protein.

XX Sequence 432 BP; 87 A; 106 C; 139 G; 100 T; 0 other;

QY Query Match 100.0%; Score 15; DB 24; Length 432;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 159 ACCTACGGCATGCAC 173

RESULT 7

ABA04563

ID ABA04563 standard; DNA; 432 BP.

XX ABA04563;

AC 15-FEB-2002 (first entry)

DT Human coding sequence, SEQ ID 63.

XX Human; cytostatic; antiinflammatory; antianaemic; vasotropic; disorder;

KW antibody; signal transduction; cancer; inflammation; hormonal

KW leukaemia; lymphoma; aplastic anaemia; skeletal malformation; ds.

XX Homo sapiens.

OS Homo sapiens.

XX Key Location/Qualifiers

FT CDS 12..419

FT /*tag= a

FT /partial

FT /product= "Human protein"

FT /note= "No stop codon given"

PN WO200179494-A1.

PD 25-OCT-2001.

PF 17-APR-2001; 2001WO-JP03288.

PR 17-APR-2000; 2000JP-0115246.

PR 20-OCT-2000; 2000JP-0321821.

PR 20-OCT-2000; 2000JP-0321822.

PR 12-MAR-2001; 2001WO-JP01912.

PA (CHUS) CHUGAI SEIYAKU KK.

PI Fukushima N, Tsuchiya M, Oheda M, Uno S, Kikuchi Y, Ohtomo T;

DR WPI; 2002-066368/09.

DR P-PSDB; AAM47642.

PT Antibodies for treatment of diseases associated with cell

PT proliferation, hormonal disorders and cytokines comprise agonist

PT activity to signal transduction across cell membranes

PS Example 7; Page 159-160; 173pp; Japanese.

XX The present invention relates to modified antibodies. The antibodies

CC contain two or more H chain V domains and two or more L chain V domains

CC of a monoclonal antibody (Mab) which is capable of transmitting a signal

CC across the cell membrane by cross-linking a cell surface molecule, where

CC the antibodies can serve as signal transduction agonists. The antibodies

CC are useful for treatment and prevention of a broad range of disorders in

CC which signal transduction is implicated, such as cancer, inflammation,

CC hormonal disorders, leukaemia, malignant lymphoma, aplastic anaemia, and

CC skeletal malformations. The present sequence was used to illustrate the

CC present invention.

XX Sequence 432 BP; 87 A; 106 C; 139 G; 100 T; 0 other;

QY Query Match 100.0%; Score 15; DB 24; Length 432;

Best Local Similarity 100.0%; Pred. No. 58;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 159 ACCTACGGCATGCAC 173

RESULT 8

AAAX17988

ID AAAX17988 standard; DNA; 717 BP.

AC AAAX17988;

XX 11-MAY-1999 (first entry)

```

XX Anti-Mpl scFv 12B5 coding sequence.
DE
XX Variant; antibody; heavy chain; light chain; immunoadhesin; immunoassay;
KW diagnosis; cancer; primer; PCR; amplification; dicistronic; ss.
XX
OS Synthetic.
XX
PN WO9850431-A2.
XX
PD 12-NOV-1998.
XX
XX 30-APR-1998; 98WO-US08762.
XX
XX 24-JUN-1997; 97US-0050661.
PR 02-MAY-1997; 97US-0850058.
XX
XX (GETH ) GENENTECH INC.
XX
XX Arathoon R, Carter PJ, Merchant AM, Presta LG;
PI
XX WPI; 1999-070091/06.
DR
XX
XX Selective preparation of multispecific antibodies - with
PT heteromultimeric heavy chain and common light chain components,
PT useful for, e.g. in vivo diagnosis of cancer
PT
XX
PS Example 4; Page -: 69pp; English.
XX
XX This sequence represents the coding sequence of the anti-Mpl scFv
CC antibody 12B5. The sequence encoding the chain is generated by a new
CC method for preparing a multispecific Ab comprising a first polypeptide
CC (PP) and at least 1 extra PP, where: (i) the first PP comprises a
CC multimerisation domain (MD) forming an interface positioned to interact
CC with an interface of a MD of the extra PP; and (ii) the first and extra
CC PPs each have a binding domain, which comprises a heavy chain and a light
CC chain, where the variable light chains of the first and extra PPs
CC comprise a common sequence. The method comprises: (a) culturing a host
CC cell comprising nucleic acid encoding the first PP and extra PP, and
CC the variable light chain, such that the nucleic acid is expressed; and
CC (b) recovering the multispecific Ab from the culture. The method prepares
CC heteromultimeric PPs, such as bispecific Abs, bispecific immunoadhesins
CC and Ab-immunoadhesin chimeras. The method allows for the enhanced
CC formation of the desired heteromultimers relative to the undesired
CC heteromultimers and homomultimers. The Abs can be used in immunoassays
CC and for the in vitro or in vivo diagnosis of various diseases, such as
CC cancer.
CC
XX
XX Sequence 717 BP; 161 A; 188 C; 215 G; 153 T; 0 other;
SQ
Query Match 100.0%; Score 15; DB 20; Length 717;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ACCTACGGCATGCAC 15
|||||
91 ACCTACGGCATGCAC 105
DB
RESULT 9
ABK71387
ID ABK71387 standard; DNA; 823 BP.
XX
AC ABK71387;
XX
DT 30-JUL-2002 (first entry)
DE
XX DNA encoding thrombopoietin agonist antibody associated protein #21.
KW Modified antibody; thrombopoietin; TPO; agonist;
KW TPO receptor; platelet reduction-associated blood disease;
KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
KW gene; ds.

```

```

XX OS Homo sapiens.
XX PN WO200233072-A1.
XX PD 25-APR-2002.
XX PF 22-OCT-2001; 2001WO-JP09259.
XX PR 20-OCT-2000; 2000JP-032182L.
XX PR 17-APR-2001; 2001WO-JP03288.
XX PR 12-SEP-2001; 2001JP-0227314.
XX PA (CHUS ) CHUGAI SEIYAKU KK.
XX PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;
XX DR WPI; 2002-383513/41.
XX DR P-PSDB; ABG35329.
XX PT Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
PT leukaemia
XX PT
XX PS Example 7; Page 190-192; 213pp; Japanese.
XX CC The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives
CC and/or remedies for platelet reduction-associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukaemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence encodes a thrombopoietin (TPO) agonist antibody associated
CC protein.
XX CC
XX SQ Sequence 823 BP; 179 A; 206 C; 248 G; 190 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 823;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
   |||
Db 159 ACCTACGGCATGCAC 173

RESULT 10
ABL56609/c
ID ABL56609 standard; cDNA; 2908 BP.
XX ABL56609;
XX AC
XX 30-JUL-2002 (first entry)
XX DT
XX DE Nucleotide sequence of human dehydrogenase 12.
XX KW Human; dehydrogenase 12; cancer; HIV infection; enzyme; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FH CDS 937..1254
FT /*tag= a
FT /product= "dehydrogenase 12"
XX PN CN1325967-A.
XX PD 12-DEC-2001.
XX PF 26-MAY-2000; 2000CN-0115887.

```

```
XX 26-MAY-2000; 2000CN-0115887.
PR
XX
PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2002-196688/26.
DR P-PSDB; ABB09889.
XX
PT Polypeptide-human dehydrogenase 12 and polynucleotide encoding it -
XX
PS Claim 6; Page 24-26 (Disclosure); 32pp; Chinese.
XX
CC The present sequence encodes human dehydrogenase 12. The polypeptide
CC is used for treating diseases such as cancer and HIV infection.
XX
SQ Sequence 2908 BP; 809 A; 539 C; 614 G; 946 T; 0 other;

Query Match          100.0%; Score 15; DB 24; Length 2908;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
   |||||
Db 1239 ACCTACGGCATGCAC 1225

RESULT 11
AAA55191
ID AAA55191 standard; DNA; 1575 BP.
XX
AC AAA55191;
XX
DT 30-AUG-2000 (first entry)
XX
DE Cenarchaeum symbiosum open reading frame nucleotide sequence SEQ ID NO:9.
XX
DE Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW characterisation; archae; therapeutic; industrial; laboratory; ds.
XX
OS Cenarchaeum symbiosum.
XX
PN WO200018909-A2.
XX
PD 06-APR-2000.
XX
PF 29-SEP-1999; 99WO-US22752.
XX
PR 29-SEP-1998; 98US-0102294.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Swanson RV, Feldman RA, Schleper C;
XX
DR WPI; 2000-293148/25.
DR P-PSDB; AAY90916.
XX
XX New nucleic acids and proteins isolated from the non-thermophilic
PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT physiology of these archae and in therapeutic, industrial or laboratory
PT techniques -
XX
PS Claim 7; Page 116-118; 210pp; English.
XX
CC AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
CC and proteins isolated from the non-thermophilic crenarchaeote
CC Cenarchaeum symbiosum. The nucleic acids and proteins identified in
CC the present invention are useful in characterising the physiology of
CC these archae and can be used in therapeutic, industrial or laboratory
CC techniques. AAA55227 to AAA55260 represent promoter sequences from
CC Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
CC probes used in examples from the present invention.
```

```
XX SQ Sequence 1575 BP; 356 A; 424 C; 548 G; 247 T; 0 other;
SQ
XX
XX Query Match          93.3%; Score 14; DB 21; Length 1575;
XX Best Local Similarity 100.0%; Pred. No. 2.3e+02;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
   |||||
Db 332 CCTACGGCATGCAC 345

RESULT 12
AAA55187
ID AAA55187 standard; DNA; 42432 BP.
XX
AC AAA55187;
XX
DT 30-AUG-2000 (first entry)
XX
DE Cenarchaeum symbiosum nucleotide sequence variant B SEQ ID NO:2.
XX
DE Cenarchaeum symbiosum; non-thermophilic; crenarchaeote; physiology;
KW characterisation; archae; therapeutic; industrial; laboratory; ds.
XX
OS Cenarchaeum symbiosum.
XX
PN WO200018909-A2.
XX
PD 06-APR-2000.
XX
PF 29-SEP-1999; 99WO-US22752.
XX
PR 29-SEP-1998; 98US-0102294.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Swanson RV, Feldman RA, Schleper C;
XX
DR WPI; 2000-293148/25.
XX
XX New nucleic acids and proteins isolated from the non-thermophilic
PT crenarchaeote Cenarchaeum symbiosum, useful in characterizing the
PT physiology of these archae and in therapeutic, industrial or laboratory
PT techniques -
XX
PS Claim 1; Page 75-87; 210pp; English.
XX
XX AAA55186 to AAA55226 and AAY90913 to AAY90951 represent nucleic acids
XX and proteins isolated from the non-thermophilic crenarchaeote
XX Cenarchaeum symbiosum. The nucleic acids and proteins identified in
XX the present invention are useful in characterising the physiology of
XX these archae and can be used in therapeutic, industrial or laboratory
XX techniques. AAA55227 to AAA55260 represent promoter sequences from
XX Cenarchaeum symbiosum. AAA55261 to AAA55269 represent PCR primers and
XX probes used in examples from the present invention.
XX
SQ Sequence 42432 BP; 8792 A; 12248 C; 12606 G; 8786 T; 0 other;

Query Match          93.3%; Score 14; DB 21; Length 42432;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
   |||||
Db 13377 CCTACGGCATGCAC 13390

RESULT 13
ABN57826
ID ABN57826 standard; DNA; 65 BP.
XX
AC ABN57826;
```


XX 15-JUL-2002 (first entry)
DT
XX
DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:30574.
XX
KW Human; mouse; rat; splice transcript; detection; RNA transcript;
KW splice variant; transcriptome; oligonucleotide library; ss.
XX
OS Mus musculus.
XX
PN WO200210449-A2.
XX
PD 07-FEB-2002.
XX
PF 20-JUL-2001; 2001WO-IB01903.
XX
PR 28-JUL-2000; 2000US-221607P.
PR 02-MAY-2001; 2001US-287724P.
XX
PA (COMP-) COMPUGEN INC.
XX
PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
XX WPI; 2002-257383/30.
DR
XX
PT New oligonucleotide libraries comprising oligonucleotides which
PT selectively hybridize to mRNAs transcribed from a transcription unit of
PT a genome, useful for detecting tissue-, pathology-, and
PT developmental-specific genes
XX
XX Example 1; SEQ ID 30574; 47pp; English.
XX
CC The present invention describes oligonucleotide libraries for detecting
CC messenger RNAs that populate a (sub-)transcriptome, where the
CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
CC transcription units that populate a genome. The library comprises
CC several oligonucleotides, each capable of hybridising selectively to a
CC set of messenger RNAs transcribed from a given transcription unit of
CC the genome, which encodes one or more messenger RNA splice variants.
CC The oligonucleotide libraries are useful for detecting mRNAs from a
CC biological sample, in expression profiling studies, in qualitatively or
CC quantitatively characterising the corresponding transcriptome, and in
CC detecting RNA transcripts and splice variants of human or animal
CC transcriptomes. The libraries may also be used as specialised mini
CC libraries to detect transcripts of a sub-transcriptome under a
CC particular biological or pathological state, and so allowing the
CC detection of tissue- and pathology-specific genes such as those genes
CC only expressed in specific tissue under a specific pathological
CC condition; to detect developmental specific genes; and to detect RNA
CC transcripts and splice variants of a transcriptome of a patient suffering
CC from a particular disorder. ABN27253 to ABN59589 represent
CC oligonucleotide sequences from rats, humans and mice, which are used in
CC the exemplification of the present invention.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 65 BP; 17 A; 20 C; 13 G; 15 T; 0 other;

Query Match 89.3%; Score 13.4; DB 24; Length 65;
Best Local Similarity 93.3%; Pred. No. 4.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
Db 14 ACCTAGGGCATGCAC 28

RESULT 14
ABL71855/C
ID ABL71855 standard; cDNA; 292 BP.
XX
AC ABL71855;

XX 14-MAY-2002 (first entry)
DT
XX
DE Corn tassel-derived polynucleotide (cdps) SEQ ID NO:1229.
XX
KW Corn; corn tassel-derived polynucleotide; cdps; hybrid breeding; CDPs;
KW inheritance; characteristic; growth; development; disease resistance;
KW environmental adaptability; quality; yield; molecular marker;
KW multigene trait; plant breeding; corn tassel; gene; ss.
XX
OS Zea mays.
XX
PN US2001051335-A1.
XX
PD 13-DEC-2001.
XX
PF 16-APR-1999; 99US-0294093.
XX
PR 21-APR-1998; 98US-082567P.
XX
PA (LALG/) LALGUDI R V.
PA (ITOL/) ITO L Y.
PA (SHER/) SHERMAN B K.
XX
XX Lalgudi RV, Ito LY, Sherman BK;
PI
XX
XX WPI; 2002-163647/21.
DR
XX
PT Novel purified corn tassel-derived polynucleotide useful for
PT determining altered gene expression, to recover regulatory elements and
PT to follow inheritance of desirable characteristics through hybrid
PT breeding programs
XX
XX
PS Claim 1; SEQ ID 1229; 201pp; English.
XX
CC The present sequence describes a purified corn tassel-derived
CC polynucleotide sequence (cdps) comprising a nucleic acid sequence
CC selected from those given in ABL70627 to ABL76833. The cdps sequences
CC encode corn tassel-derived polypeptides (CDPs). The cdps sequences (I)
CC can be used for determining altered gene expression, to recover
CC regulatory elements and to follow inheritance of desirable
CC characteristics through hybrid breeding programs. (I) are also useful
CC in the evaluation, and alteration of desired characteristics associated
CC with growth and development, disease resistance, environmental
CC adaptability, quality and yield, and as molecular markers for studying
CC inheritance of multigene traits in a plant breeding program. (I) can be
CC used to produce a tassel-specific profile of gene transcription, a
CC transcript image, to clone regulatory elements for use in transformation
CC vectors, to express a polypeptide, to identify, isolate or extend
CC identical or related corn tassel nucleic acid sequences from DNA
CC libraries, in nucleic acid hybridisation or amplification technologies,
CC as query sequences to determine homology of known sequences, as probe
CC for use in Southern or Northern hybridisation, and to identify the
CC presence of and/or to determine the degree of similarity between two
CC (or more) nucleic acid sequences.
XX
SQ Sequence 292 BP; 78 A; 56 C; 85 G; 67 T; 6 other;

Query Match 89.3%; Score 13.4; DB 24; Length 292;
Best Local Similarity 93.3%; Pred. No. 4.6e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
Db 176 ACCTAAGGCATGCAC 162

RESULT 15
AAH68618
ID AAH68618 standard; DNA; 372 BP.
XX
AC AAH68618;
XX

```

DT      14-SEP-2001      (first entry)
XX
DE      Human anti-Rh(D) chain C04 nucleotide sequence.
XX
KW      Human; anti-Rh(D) antibody; Rh(D) binding protein; immunostimulant;
XX      red blood cell; Rh phenotype; diagnosis; therapeutic; ds.
OS      Homo sapiens.
PN      US6255455-B1.
XX
PD      03-JUL-2001.
XX
PF      29-JAN-1999;      99US-0240274.
XX
PR      11-OCT-1996;      96US-0028550.
PR      10-APR-1998;      98US-0081380.
PR      27-JUN-1997;      97US-0884045.
XX
PA      (TYPE-) UNITV PENNSYLVANIA.
XX
PI      Siegel DL;
XX
DR      WPI; 2001-388931/41.
DR      P-PSDB; AAG93561.
XX
PT      New isolated protein, preferably a human anti-Rh(D) antibody for use in
PT      diagnostics requiring a human instead of an animal antibody and in
PT      therapeutic medicine -
XX
XX
XX      Example 2; Column 47; 162pp; English.
XX
XX      The present invention describes an isolated Rh(D) binding protein,
XX      preferably a human antibody, (I) having an amino acid sequence comprising
XX      one of the sequences (S) given in AAG93558 to AAG93669. (I) has
XX      immunostimulant activity, and can be used as an immune system stimulant.
XX      (I) can be used in diagnostic and therapeutic medicine. The antibodies
XX      are used in diagnostics that require human antibodies instead of animal
XX      antibodies, such as determine the Rh phenotype of human red blood cells.
XX      AAH68615 to AAH68726 represent the nucleotide sequence which encode
XX      AAG93558 to AAG93669. AAG93670 to AAG93697 represent anti-Rh(D) heavy
XX      chain CDR3 amino acid sequences which are given in the exemplification
XX      of the present invention.
XX
SQ      Sequence 372 BP; 86 A; 92 C; 108 G; 86 T; 0 other;

Query Match      89.3%;      Score 13.4;      DB 22;      Length 372;
Best Local Similarity      93.3%;      Pred. No. 4.7e+02;
Matches      14;      Conservative      0;      Mismatches      1;      Indels      0;      Gaps      0;

1 ACCTACGGCATGCAC 15
||||| |||||||||
91 ACCTATGGCATGCAC 105

```

```
Search completed: June 26, 2003, 03:11:48
Job time : 88 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-25

Perfect score: 15
Sequence: 1 acctacgcatgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	4	US-08-918-148-25 Sequence 25, Appl
2	13.4	89.3	372	4	US-09-240-274-73 Sequence 73, Appl
3	13.4	89.3	375	4	US-09-240-274-78 Sequence 78, Appl
4	13.4	89.3	375	4	US-09-240-274-93 Sequence 93, Appl
5	13.4	89.3	2348	4	US-09-388-743-5 Sequence 5, Appli
6	13.4	89.3	3735	4	US-08-975-762-43 Sequence 43, Appl
7	13.4	89.3	3735	4	US-09-295-028-43 Sequence 43, Appl
8	13.4	89.3	3735	4	US-09-106-582-43 Sequence 57, Appl
9	13.4	89.3	87563	4	US-09-453-702B-57 Sequence 6, Appli
10	12.4	82.7	24	4	US-09-673-018-6 Sequence 18, Appl
11	12.4	82.7	191	5	PCT-US95-09114-18 Sequence 15, Appl
12	12.4	82.7	265	4	US-08-990-823-15 Sequence 18, Appl
13	12.4	82.7	302	3	US-08-964-268-18 Sequence 182, App
14	12.4	82.7	381	4	US-09-240-274-182 Sequence 45, Appl
15	12.4	82.7	717	4	US-09-513-783A-45 Sequence 3, Appli
16	12.4	82.7	720	3	US-09-172-063-11 Sequence 11, Appl
17	12.4	82.7	720	4	US-09-316-919-12 Sequence 12, Appl
18	12.4	82.7	818	4	US-09-221-017B-900 Sequence 900, App
19	12.4	82.7	850	3	US-09-062-102-2 Sequence 2, Appli
20	12.4	82.7	850	4	US-09-364-946-2 Sequence 29, Appl
21	12.4	82.7	972	3	US-09-172-063-29 Sequence 176, App
22	12.4	82.7	1001	4	US-09-641-638-176 Sequence 177, App
23	12.4	82.7	1001	4	US-09-641-638-177 Sequence 178, App
24	12.4	82.7	1001	4	US-09-641-638-178 Sequence 5, Appli
25	12.4	82.7	1095	4	US-09-085-305-5 Sequence 7, Appli
26	12.4	82.7	1155	2	US-08-387-942C-7
27	12.4	82.7	1155	2	US-08-387-942C-7

28	12.4	82.7	1155	2	US-08-387-942C-18	Sequence 18, Appl
29	12.4	82.7	1176	2	US-08-387-942C-17	Sequence 17, Appl
30	12.4	82.7	1380	4	US-09-513-783A-169	Sequence 169, App
31	12.4	82.7	1419	4	US-09-516-914-4	Sequence 4, Appli
32	12.4	82.7	1515	4	US-09-080-625-6	Sequence 6, Appli
33	12.4	82.7	1515	4	US-09-695-782-6	Sequence 6, Appli
34	12.4	82.7	1521	1	US-08-726-136-20	Sequence 20, Appl
35	12.4	82.7	1521	3	US-09-103-434-20	Sequence 20, Appl
36	12.4	82.7	1521	4	US-09-687-594-20	Sequence 20, Appl
37	12.4	82.7	1586	4	US-09-673-018-1	Sequence 1, Appli
38	12.4	82.7	1650	1	US-08-459-100A-1	Sequence 1, Appli
39	12.4	82.7	1650	5	PCT-US94-09589-1	Sequence 1, Appli
40	12.4	82.7	1770	4	US-09-513-783A-1	Sequence 1, Appli
41	12.4	82.7	1929	2	US-08-818-253-1	Sequence 1, Appli
42	12.4	82.7	1929	2	US-08-818-253-5	Sequence 5, Appli
43	12.4	82.7	1929	4	US-08-818-252-1	Sequence 1, Appli
44	12.4	82.7	1929	4	US-08-818-252-5	Sequence 5, Appli
45	12.4	82.7	1959	2	US-08-818-253-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-08-918-148-25
Sequence 25, Application US/08918148A
Patent No. 6342220
GENERAL INFORMATION:
APPLICANT: Adams, Camellia
APPLICANT: W.
APPLICANT: Carter, Paul J.
APPLICANT: Fendly, Brian M.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Agonist Antibodies
FILE REFERENCE: P0979
CURRENT APPLICATION NUMBER: US/08/918,148A
CURRENT FILING DATE: 1997-08-25
NUMBER OF SEQ ID NOS: 79
SEQ ID NO 25
LENGTH: 15
TYPE: DNA
ORGANISM: artificial
FEATURE:
NAME/KEY: 10D10scFv, 12B5scFv VH CDR1
LOCATION: 1-15
OTHER INFORMATION:
US-08-918-148-25

Query Match 100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. NO. 10;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
Db 1 ACCTACGGCATGCAC 15

RESULT 2
US-09-240-274-73
Sequence 73, Application US/09240274
Patent No. 6255455
GENERAL INFORMATION:
APPLICANT: Siegel, Donald L.
TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
TITLE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
FILE REFERENCE: 09596-4202
CURRENT APPLICATION NUMBER: US/09/240,274
CURRENT FILING DATE: 1999-01-29
EARLIER APPLICATION NUMBER: 60/081,380
EARLIER FILING DATE: 1998-04-10
EARLIER APPLICATION NUMBER: 60/028,550
EARLIER FILING DATE: 1996-10-11
NUMBER OF SEQ ID NOS: 224

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain C04
US-09-240-274-73

```

```

Query Match      89.3%; Score 13.4; DB 4; Length 372;
Best Local Similarity 93.3%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||| ||||| |||||
Db       91 ACCTATGGCATGCAC 105

```

RESULT 3

```

US-09-240-274-78
; Sequence 78, Application US/09240274
; Patent No. 6255455
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/240,274
; CURRENT FILING DATE: 1999-01-29
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/081,380
; EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain D03
US-09-240-274-78

```

```

Query Match      89.3%; Score 13.4; DB 4; Length 375;
Best Local Similarity 93.3%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||| ||||| |||||
Db       91 ACCTATGGCATGCAC 105

```

RESULT 4

```

US-09-240-274-93
; Sequence 93, Application US/09240274
; Patent No. 6255455
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/240,274
; CURRENT FILING DATE: 1999-01-29
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/028,550
; EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```

; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain D20
US-09-240-274-93

```

```

Query Match      89.3%; Score 13.4; DB 4; Length 375;
Best Local Similarity 93.3%; Pred. No. 87;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||| ||||| |||||
Db       91 ACCTATGGCATGCAC 105

```

RESULT 5

```

US-09-388-743-5/c
; Sequence 5, Application US/09388743
; Patent No. 6423886
; GENERAL INFORMATION:
; APPLICANT: Singletary, George
; APPLICANT: Zhou, Ian
; TITLE OF INVENTION: No. 6423886el Starch Synthase Polynucleotides and Their
; FILE REFERENCE: 1144
; CURRENT APPLICATION NUMBER: US/09/388,743
; CURRENT FILING DATE: 1999-09-02
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 2348
; TYPE: DNA
; ORGANISM: Curcuma zedoaria
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (36)...(2105)
US-09-388-743-5

```

```

Query Match      89.3%; Score 13.4; DB 4; Length 2348;
Best Local Similarity 93.3%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||| ||||| |||||
Db       1868 ACCTACTGCATGCAC 1854

```

RESULT 6

```

US-08-975-762-43
; Sequence 43, Application US/08975762
; Patent No. 6207169
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,762
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.

```

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 3735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-975-762-43

Query Match 89.3%; Score 13.4; DB 4; Length 3735;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
||||| |||||||
Db 1390 ACCTACTGCATGCAC 1404

RESULT 7
US-09-295-028-43
Sequence 43, Application US/09295028
Patent No. 6277381
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF EHRLICHIA INFECTION
FILE REFERENCE: 210121.439C4
CURRENT APPLICATION NUMBER: US/09/295,028
CURRENT FILING DATE: 1999-04-20
NUMBER OF SEQ ID NOS: 85
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 43
LENGTH: 3735
TYPE: DNA
ORGANISM: Ehrlichia sp.
US-09-295-028-43

Query Match 89.3%; Score 13.4; DB 4; Length 3735;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
||||| |||||||
Db 1390 ACCTACTGCATGCAC 1404

RESULT 8
US-09-106-582-43
Sequence 43, Application US/09106582
Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 3735 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-106-582-43

Query Match 89.3%; Score 13.4; DB 4; Length 3735;
Best Local Similarity 93.3%; Pred. No. 95;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
||||| |||||||
Db 1390 ACCTACTGCATGCAC 1404

RESULT 9
US-09-453-702B-57/c
Sequence 57, Application US/09453702B
Patent No. 6365723
GENERAL INFORMATION:
APPLICANT: Blatner, Frederick R.
Burland, Valerie
Perna, Nicole T.
Plunkett, Guy
Welch, Rod
TITLE OF INVENTION: NO. 6365723e1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pluckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch. 1.44Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 87563

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 57:
US-09-453-702B-57

Query Match 89.3%; Score 13.4; DB 4; Length 87563;
Best Local Similarity 93.3%; Pred. No. 1.1e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
Db 41493 ACCTACGGCATGCAC 41479

RESULT 10

US-09-673-018-6/c
Sequence 6, Application US/09673018
Patent No. 6461842
GENERAL INFORMATION:
APPLICANT: MATSUDA, Hideyuki
APPLICANT: KAWAMUKAI, Makoto
APPLICANT: YAJIMA, Kazuyoshi
APPLICANT: IKENAKA, Yasuhiko
APPLICANT: NISHI, Kenichi
APPLICANT: HASEGAWA, Junzo
APPLICANT: TAKAHASHI, Satomi
TITLE OF INVENTION: METHOD FOR PREPARING COENZYME Q10
FILE REFERENCE: 2000-1379A/LC/00177
CURRENT APPLICATION NUMBER: US/09/673, 018
CURRENT FILING DATE: 2000-10-10
PRIOR APPLICATION NUMBER: JP 11/32657
PRIOR FILING DATE: 1999-02-10
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 24
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: DNA Primers
US-09-673-018-6

Query Match 82.7%; Score 12.4; DB 4; Length 24;
Best Local Similarity 92.9%; Pred. No. 2.8e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
Db 19 CCTACGGCATGAAC 6

RESULT 11

PCT-US95-09114-18
Sequence 18, Application PC/TUS9509114
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Antibodies that Bind a Conformationally
TITLE OF INVENTION: Altered CD4 Molecule Induced Upon Binding of
TITLE OF INVENTION: Human Immunodeficiency Virus
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09114
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/277, 080
FILING DATE: 19-Jul-94
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/305, 903
FILING DATE: 13-Sep-94
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Deconti, Giulio A., Jr.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: BIZ-013CPPC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US95-09114-18

Query Match 82.7%; Score 12.4; DB 5; Length 191;
Best Local Similarity 92.9%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCA 14
Db 91 ACCTACGGCATGCA 104

RESULT 12

US-08-990-823-15/c
Sequence 15, Application US/08990823D
Patent No. 6228371
GENERAL INFORMATION:
APPLICANT: Nano, Francis
TITLE OF INVENTION: Mycobacterium Tuberculosis DNA Sequences Encoding
TITLE OF INVENTION: Immunostimulatory Peptides
FILE REFERENCE: 49086
CURRENT APPLICATION NUMBER: US/08/990, 823D
CURRENT FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: US 96/10375
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000, 254
EARLIER FILING DATE: 1995-06-15
NUMBER OF SEQ ID NOS: 113
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 15
LENGTH: 265
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
NAME/KEY: Modified base
OTHER INFORMATION: n represents a or g or c or t/u
US-08-990-823-15

Query Match 82.7%; Score 12.4; DB 4; Length 265;
Best Local Similarity 92.9%; Pred. No. 3e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
Db 33 CGTACGGCATGCAC 20

```
RESULT 13
US-08-964-268-18
; Sequence 18, Application US/08964268
; Patent No. 6114503
; GENERAL INFORMATION:
; APPLICANT: WEI, YING-FEI
; APPLICANT: RUBEN, STEVEN M
; APPLICANT: SANCAR, AZIZ
; APPLICANT: HSU, SHIAO-WEN D
; APPLICANT: KAZANTSEV, ALEKSEY G
; TITLE OF INVENTION: HUMAN BLUE-LIGHT PHOTORECEPTOR hCRY2
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,268
; FILING DATE: HERewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,189
; FILING DATE: 04-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: STEEFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0750001/EKS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-964-268-18

Query Match      82.7%; Score 12.4; DB 3; Length 302;
Best Local Similarity 86.7%; Pred. No. 3.1e+02;
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCAC 15
      |||||||  |||  |
Db      126 ACCTACGGCGTGCNC 140

RESULT 14
US-09-240-274-182
; Sequence 182, Application US/09240274
; Patent No. 6255455
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; TITLE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/240,274
; CURRENT FILING DATE: 1999-01-29
; EARLIER APPLICATION NUMBER: 60/081,380
; EARLIER FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/028,550
; EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 182
```

```
; LENGTH: 381
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) antibody clone SH10
US-09-240-274-182

Query Match      82.7%; Score 12.4; DB 4; Length 381;
Best Local Similarity 92.9%; Pred. No. 3.1e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 CCTACGGCATGCAC 15
      |||||||  |||  |
Db      332 CCTACGGCATGCAC 345

RESULT 15
US-09-513-783A-45
; Sequence 45, Application US/09513783A
; Patent No. 6416959
; GENERAL INFORMATION:
; APPLICANT: Giuliano, Kenneth A.
; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-L1
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 45
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(717)
; OTHER INFORMATION: Description of Artificial Sequence: EGFP
US-09-513-783A-45

Query Match      82.7%; Score 12.4; DB 4; Length 717;
Best Local Similarity 92.9%; Pred. No. 3.2e+02;
Matches 13; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACCTACGGCATGCA 14
      |||||||  |||  |
Db      196 ACCTACGGCGTGCA 209

Search completed: June 26, 2003, 04:52:26
Job time : 30.2 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-25
Perfect score: 15
Sequence: 1 acctacgcatgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	14	93.3	1575	9	US-10-027-806-9	Sequence 9, Appli
2	14	93.3	1575	9	US-10-034-623-9	Sequence 9, Appli
3	14	93.3	1575	9	US-10-027-801-9	Sequence 9, Appli
4	14	93.3	42432	9	US-10-027-806-2	Sequence 2, Appli
5	14	93.3	42432	9	US-10-034-623-2	Sequence 2, Appli
6	14	93.3	42432	9	US-10-027-801-2	Sequence 2, Appli
7	13.4	89.3	292	10	US-09-294-093B-1229	Sequence 1229, Ap
8	13.4	89.3	372	9	US-09-848-798-73	Sequence 73, Appl
9	13.4	89.3	375	9	US-09-848-798-78	Sequence 78, Appl
10	13.4	89.3	375	9	US-09-848-798-93	Sequence 93, Appl
11	13.4	89.3	375	9	US-10-172-317-1	Sequence 1, Appli
12	13.4	89.3	714	9	US-09-822-846-612	Sequence 612, App
13	13.4	89.3	933	10	US-09-838-561-6	Sequence 6, Appli
14	13.4	89.3	933	10	US-09-816-760-6	Sequence 6, Appli
15	13.4	89.3	1157	10	US-09-820-893-31	Sequence 31, Appl
16	13.4	89.3	1379	10	US-09-838-561-4	Sequence 4, Appli
17	13.4	89.3	1379	10	US-09-816-760-4	Sequence 4, Appli
18	13.4	89.3	1743	9	US-09-738-626-1827	Sequence 1827, Ap
19	13.4	89.3	3735	10	US-09-159-469-43	Sequence 43, Appl

20	13.4	89.3	3735	10	US-09-798-042-43	Sequence 43, Appl
21	13.4	89.3	3735	10	US-09-798-042-88	Sequence 88, Appl
22	13.4	89.3	3735	10	US-09-798-042-96	Sequence 96, Appl
23	13.4	89.3	87563	9	US-10-114-170-57	Sequence 57, Appl
24	13.4	89.3	3309400	9	US-09-738-626-1	Sequence 1, Appli
25	13	86.7	477	9	US-09-738-626-1449	Sequence 1449, Ap
26	13	86.7	498	9	US-09-918-995-11630	Sequence 11630, A
27	13	86.7	1311	10	US-09-925-297-354	Sequence 354, App
28	12.4	82.7	53	10	US-09-962-628B-19	Sequence 19, Appl
29	12.4	82.7	53	10	US-09-962-628B-20	Sequence 20, Appl
30	12.4	82.7	53	10	US-09-962-628B-22	Sequence 22, Appl
31	12.4	82.7	144	10	US-09-294-093B-3458	Sequence 3458, Ap
32	12.4	82.7	157	10	US-09-864-761-30169	Sequence 30169, A
33	12.4	82.7	265	9	US-09-996-634-15	Sequence 15, Appl
34	12.4	82.7	265	9	US-09-997-181-15	Sequence 15, Appl
35	12.4	82.7	265	9	US-09-997-182-15	Sequence 15, Appl
36	12.4	82.7	293	10	US-09-864-761-24230	Sequence 24230, A
37	12.4	82.7	378	9	US-09-835-976B-69	Sequence 69, Appl
38	12.4	82.7	381	9	US-09-848-798-182	Sequence 182, App
39	12.4	82.7	388	9	US-09-954-531-724	Sequence 724, App
40	12.4	82.7	462	10	US-09-770-444-404	Sequence 404, App
41	12.4	82.7	536	10	US-09-974-300-955	Sequence 955, App
42	12.4	82.7	558	10	US-09-864-761-7525	Sequence 7525, App
43	12.4	82.7	575	10	US-09-864-761-13641	Sequence 13641, A
44	12.4	82.7	717	9	US-09-989-025A-1	Sequence 1, Appli
45	12.4	82.7	717	9	US-10-100-957A-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-10-027-806-9
; Sequence 9, Application US/10027806
; Patent No. US20020160476A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCCORP.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; US-10-027-806-9

Query Match 93.3%; Score 14; DB 9; Length 1575;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
Db 332 CCTACGGCATGCAC 345

RESULT 2
US-10-034-623-9
; Sequence 9, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa

```
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCorp.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; US-10-034-623-9
```

```
Query Match          93.3%; Score 14; DB 9; Length 1575;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 CCTACGGCATGCAC 15
        |||
Db      332 CCTACGGCATGCAC 345
```

RESULT 3

```
; US-10-027-801-9
; Sequence 9, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCorp.002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1575)
; US-10-027-801-9
```

```
Query Match          93.3%; Score 14; DB 9; Length 1575;
Best Local Similarity 100.0%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 CCTACGGCATGCAC 15
        |||
Db      332 CCTACGGCATGCAC 345
```

RESULT 4

```
; US-10-027-806-2
; Sequence 2, Application US/10027806
; Patent No. US20020160476A1.
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCorp.002A
; CURRENT APPLICATION NUMBER: US/10/027,806
```

```
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
; US-10-027-806-2
```

```
Query Match          93.3%; Score 14; DB 9; Length 42432;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 CCTACGGCATGCAC 15
        |||
Db      13377 CCTACGGCATGCAC 13390
```

RESULT 5

```
; US-10-034-623-2
; Sequence 2, Application US/10034623
; Publication No. US20020198365A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCorp.002A
; CURRENT APPLICATION NUMBER: US/10/034,623
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 60/102,294
; PRIOR FILING DATE: 1998-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS
```



```
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
; US-10-034-623-2
```

```
Query Match      93.3%; Score 14; DB 9; Length 42432;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 CCTACGGCATGCAC 15
DB      13377 CCTACGGCATGCAC 13390
```

RESULT 6

```
US-10-027-801-2
; Sequence 2, Application US/10027801
; Publication No. US20030054364A1
; GENERAL INFORMATION:
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Feldman, Robert A.
; APPLICANT: Schleper, Christa
; TITLE OF INVENTION: NUCLEIC ACIDS AND PROTEINS FROM CENARCHAEUM SYMBIOSUM
; FILE REFERENCE: DCOB-002A
; CURRENT APPLICATION NUMBER: US/10/027,801
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/408,020
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29
; NUMBER OF SEQ ID NOS: 123
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 42432
; TYPE: DNA
; ORGANISM: Cenarchaeum symbiosum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)...(10421)
; NAME/KEY: CDS
; LOCATION: (10625)...(11434)
; NAME/KEY: CDS
; LOCATION: (11478)...(13046)
; NAME/KEY: CDS
; LOCATION: (13046)...(14620)
; NAME/KEY: CDS
; LOCATION: (23558)...(24862)
; NAME/KEY: CDS
; LOCATION: (24913)...(25728)
; NAME/KEY: CDS
; LOCATION: (26504)...(26881)
; NAME/KEY: CDS
; LOCATION: (29655)...(30491)
; NAME/KEY: CDS
; LOCATION: (34559)...(36067)
; NAME/KEY: CDS
; LOCATION: (37002)...(37403)
; NAME/KEY: CDS
; LOCATION: (37404)...(38282)
```

```
; NAME/KEY: CDS
; LOCATION: (39454)...(40572)
; US-10-027-801-2
```

```
Query Match      93.3%; Score 14; DB 9; Length 42432;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      2 CCTACGGCATGCAC 15
DB      13377 CCTACGGCATGCAC 13390
```

RESULT 7

```
US-09-294-093B-1229/C
; Sequence 1229, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: PERL Program
; SEQ ID NO 1229
; LENGTH: 292
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700343977H1
; NAME/KEY: unsure
; LOCATION: 52, 62, 68, 83, 90, 145
; OTHER INFORMATION: a, t, c, g, or other
; US-09-294-093B-1229
```

```
Query Match      89.3%; Score 13.4; DB 10; Length 292;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 ACCTACGGCATGCAC 15
DB      176 ACCTAAGCATGCAC 162
```

RESULT 8

```
US-09-848-798-73
; Sequence 73, Application US/09848798
; Publication No. US20030040605A1
; GENERAL INFORMATION:
; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; FILE REFERENCE: 09596-4202
; CURRENT APPLICATION NUMBER: US/09/848,798
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/240,274
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/028,550
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 73
; LENGTH: 372
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain C04
```

US-09-848-798-73

Query Match 89.3%; Score 13.4; DB 9; Length 372;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
||||| |||||||||
Db 91 ACCTATGGCATGCAC 105

RESULT 9

US-09-848-798-78
; Sequence 78, Application US/09848798
; Publication No. US20030040605A1
; GENERAL INFORMATION:

; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; TITLE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
; FILE REFERENCE: 09596-42U2
; CURRENT APPLICATION NUMBER: US/09/848,798
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/240,274
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/028,550
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 78
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain D03
US-09-848-798-78

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
||||| |||||||||
Db 91 ACCTATGGCATGCAC 105

RESULT 10

US-09-848-798-93
; Sequence 93, Application US/09848798
; Publication No. US20030040605A1
; GENERAL INFORMATION:

; APPLICANT: Siegel, Donald L.
; TITLE OF INVENTION: Rh(D)-BINDING PROTEINS AND MAGNETICALLY ACTIVATED CELL
; TITLE OF INVENTION: SORTING METHOD FOR PRODUCTION THEREOF
; FILE REFERENCE: 09596-42U2
; CURRENT APPLICATION NUMBER: US/09/848,798
; CURRENT FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/240,274
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/028,550
; PRIOR FILING DATE: EARLIER FILING DATE: 1996-10-11
; NUMBER OF SEQ ID NOS: 224
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 93
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: anti-Rh(D) chain D20
US-09-848-798-93

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
||||| |||||||||
Db 91 ACCTATGGCATGCAC 105

RESULT 11

US-10-172-317-1
; Sequence 1, Application US/10172317
; Publication No. US20030091561A1
; GENERAL INFORMATION:
; APPLICANT: van de Winkel, Jan G.J.
; APPLICANT: van Dijk, Marcus A.
; APPLICANT: Halk, Edward
; APPLICANT: Gerritsen, Arnout F.
; TITLE OF INVENTION: HUMAN MONOCLONAL ANTIBODIES TO
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR (EGFR)
; FILE REFERENCE: GMI-020
; CURRENT APPLICATION NUMBER: US/10/172,317
; CURRENT FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/298,172
; PRIOR FILING DATE: 2001-06-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 375
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-172-317-1

Query Match 89.3%; Score 13.4; DB 9; Length 375;
Best Local Similarity 93.3%; Pred. No. 2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
||||| |||||||||
Db 91 ACCTATGGCATGCAC 105

RESULT 12

US-09-822-846-612/c
; Sequence 612, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:

; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaValle, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steinger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES' ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 612
; LENGTH: 714

```

; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-612

```

```

Query Match      89.3%; Score 13.4; DB 9; Length 714;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||||| |||||
Db      219 ACCTACGGTATGCAC 205

```

RESULT 13

```

US-09-838-561-6
; Sequence 6, Application US/09838561
; Patent No. US20020042371A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Meyers, Rachel
; APPLICANT: Cook, William James
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; APPLICANT: Gimeno, Ruth
; TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL DEHYDROGENASE
; TITLE OF INVENTION: MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-134CP2
; CURRENT APPLICATION NUMBER: US/09/838,561
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: 09/816,760
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 09/634,955
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 60/192,002
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 6
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(933)
US-09-838-561-6

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 933;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||||| |||||
Db      745 ACCTACGGCGTGCAC 759

```

RESULT 14

```

US-09-816-760-6
; Sequence 6, Application US/09816760
; Patent No. US20020052032A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Meyers, Rachel
; APPLICANT: Cook, William James
; APPLICANT: Williamson, Mark
; APPLICANT: Rudolph-Owen, Laura A.
; TITLE OF INVENTION: 32142, 21481, 25964, 21686, NOVEL HUMAN DEHYDROGENASE
; TITLE OF INVENTION: MOLECULES AND USES THEREFOR
; FILE REFERENCE: MNI-134CP
; CURRENT APPLICATION NUMBER: US/09/816,760
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 09/634,955
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: 60/192,002
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 13

```

```

; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 933
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(933)
US-09-816-760-6

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 933;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||||| |||||
Db      745 ACCTACGGCGTGCAC 759

```

RESULT 15

```

US-09-820-893-31
; Sequence 31, Application US/09820893
; Patent No. US20020076705A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 31 Human Secreted Proteins
; FILE REFERENCE: P2033P1
; CURRENT APPLICATION NUMBER: US/09/820,893
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/531,119
; PRIOR FILING DATE: 2000-03-20
; PRIOR APPLICATION NUMBER: 60/102,895
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 140
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 31
; LENGTH: 1157
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-820-893-31

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 1157;
Best Local Similarity 93.3%; Pred. No. 1.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 ACCTACGGCATGCAC 15
        ||||||| |||||
Db      848 ACCTACGGCGTGCAC 862

```

Search completed: June 26, 2003, 04:57:12
Job time : 65.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-25
Perfect score: 15
Sequence: 1 acctacgcgcac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	15	100.0	354	14	R12983	R12983 yf70e11.r1
C 2	15	100.0	549	12	BF326887	BF326887 QV3-BN004
C 3	15	100.0	663	12	BG539077	BG539077 602568449
C 4	15	100.0	706	14	BM716521	BM716521 UI-E-EJ0-
C 5	15	100.0	923	12	BE882505	BE882505 601507246
C 6	15	100.0	961	14	BQ719668	BQ719668 AGENCOURT

C 7	14	93.3	268	14	T82729	T82729 Ni9011 Aspe
C 8	14	93.3	325	9	A1764493	A1764493 UI-R-Y0-a
C 9	14	93.3	369	13	B1183951	B1183951 UNL-P-FN-
C 10	14	93.3	486	14	BM728351	BM728351 UI-E-E01-
C 11	14	93.3	531	14	BQ040153	BQ040153 gd12b08.y
C 12	14	93.3	553	17	AQ657745	AQ657745 Sheared D
C 13	14	93.3	569	10	AV386289	AV386289 AV386289
C 14	14	93.3	591	17	AQ943516	AQ943516 Sheared D
C 15	14	93.3	591	17	A2103592	A2103592 RPCI-23-2
C 16	14	93.3	591	17	AQ652602	AQ652602 Sheared D
C 17	14	93.3	602	17	AQ652520	AQ652520 Sheared D
C 18	14	93.3	608	10	AV385351	AV385351 AV385351
C 19	14	93.3	616	17	AQ660685	AQ660685 Sheared D
C 20	14	93.3	674	10	AV385001	AV385001 AV385001
C 21	14	93.3	696	17	BH545455	BH545455 BOCJP01TR
C 22	14	93.3	761	17	BH696405	BH696405 BOMV16TR
C 23	14	93.3	877	12	BF028074	BF028074 601764313
C 24	14	93.3	1096	12	BG474704	BG474704 602517518
C 25	13.4	89.3	105	12	BE860369	BE860369 UI-M-A01-
C 26	13.4	89.3	105	17	BH865555	BH865555 SALK_0989
C 27	13.4	89.3	137	17	AQ646650	AQ646650 RPCI93-EC
C 28	13.4	89.3	162	10	AW984704	AW984704 PM4-HN001
C 29	13.4	89.3	200	9	AU076815	AU076815 AU076815
C 30	13.4	89.3	203	14	D30996	D30996 HUML12083 H
C 31	13.4	89.3	209	10	BE365107	BE365107 P11_24_F0
C 32	13.4	89.3	213	10	BE157960	BE157960 MR2-HT037
C 33	13.4	89.3	216	12	BF088542	BF088542 RC1-HT088
C 34	13.4	89.3	231	17	AZ311664	AZ311664 1M0027M12
C 35	13.4	89.3	248	14	BM784367	BM784367 K-EST0062
C 36	13.4	89.3	248	14	H00321	H00321 YJ23f04.S1
C 37	13.4	89.3	264	10	AV328192	AV328192 AV328192
C 38	13.4	89.3	274	14	F35823	F35823 HSPD32901 H
C 39	13.4	89.3	288	13	BM193261	BM193261 TCBAPIE50
C 40	13.4	89.3	289	10	BB489547	BB489547 BB489547
C 41	13.4	89.3	291	10	AW461362	AW461362 BP230004A
C 42	13.4	89.3	297	14	T27868	T27868 EST18962 Hu
C 43	13.4	89.3	307	9	A1470361	A1470361 tJ42f03.x
C 44	13.4	89.3	307	17	BH239762	BH239762 ATYCM26TF
C 45	13.4	89.3	308	17	FR0003712	287495 F.rubripes

ALIGNMENTS

RESULT 1
R12983/c 354 bp mRNA linear EST 12-APR-1995
LOCUS yf70e11.r1 Soares infant brain INIB Homo sapiens cDNA clone
DEFINITION IMAGE:27463 5', mRNA sequence.
ACCESSION R12983
VERSION R12983.1 GI:766059
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 354)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
,R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2106
High quality sequence stops: 175 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

TITLE JOURNAL
COMMENT

QY 1 ACCTACGGCATGCAC 15
|||||
Db 92 ACCTACGGCATGCAC 78

RESULT 4
BM716521/c 706 bp mRNA linear EST 28-FEB-2002
LOCUS
DEFINITION UI-E-EJ0-ah1-i-11-0-UI.r2 UI-E-EJ0 Homo sapiens cDNA clone
UI-E-EJ0-ah1-i-11-0-UI 5', mRNA sequence.
ACCESSION BM716521
VERSION BM716521 GI:19029779
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 706)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mscoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
CDNA Library preparation: Dr. M. Bento Soares, Univeristy of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Reverse.

FEATURES
source
1..706
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-E-EJ0-ah1-i-11-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT7T3-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
optic nerve, CCATTAGTGC; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."

BASE COUNT 183 a 130 c 164 g 228 t 1 others
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 706;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
Db 28 ACCTACGGCATGCAC 14

RESULT 5
BE882505/c 923 bp mRNA linear EST 20-OCT-2000
LOCUS
DEFINITION 601507246F1 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:3908891 5',
mRNA sequence.
ACCESSION BE882505
VERSION BE882505 GI:10331281
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 923)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9721 row: 1 column: 12
High quality sequence stop: 608.

FEATURES
source
1..923
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3908891"
/clone_lib="NIH_MGC_71"
/tissue_type="leiomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2.1 kb."

BASE COUNT 170 a 228 c 255 g 270 t
ORIGIN

Query Match 100.0%; Score 15; DB 12; Length 923;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||||
Db 478 ACCTACGGCATGCAC 464

RESULT 6
BQ719668/c 961 bp mRNA linear EST 16-JUL-2002
LOCUS
DEFINITION AGENCOURT_8304828 lupski_sympathetic_trunk Homo sapiens cDNA clone
IMAGE:6193699 5', mRNA sequence.
ACCESSION BQ719668
VERSION BQ719668 GI:21858565
KEYWORDS
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 961)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov

Tissue Procurement: Dr. James R. Lupski
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM13597 row: 1 column: 20
High quality sequence stop: 583.

FEATURES
Source
1. .961
location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6193699"
/clone_lib="Lupski_sympathetic_trunk"
/sex="male"
/tissue_type="sympathetic trunk"
/dev_stage="adult, 16 yr"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6 (Life Technologies); Site_1:
NotI; Site_2: SalI; cDNA made by oligo-dT priming.
Directionally cloned using the following adaptors:
5'-TCGACCCACGCGTCCG-3' and
5'-GACTAGTCTTAGATCGGAGCGGCCCT(15)-3'. Size selected >
1 kb for average insert length 1.9 kb. This is a primary
library, non-amplified. Library constructed by Life
Technologies and donated by J. Lupski, M.D./Ph.D. (Baylor
College of Medicine); available through Life
Technologies."

BASE COUNT
ORIGIN
187 a 274 c 248 g 252 t

Query Match 100.0%; Score 15; DB 14; Length 961;
Best Local Similarity 100.0%; Pred. No. 1.8e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCAC 15
|||||
Db 574 ACCTACGGCATGCAC 560

RESULT 7
T82729 268 bp mRNA linear EST 18-SEP-2000
LOCUS N16011 Aspergillus niger, PYES2 (XhoI-EcoRI) Aspergillus niger cDNA
DEFINITION 5' end, mRNA sequence.
T82729
T82729.1 GI:711017

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Aspergillus niger.
Aspergillus niger
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 268)
Williams, B.A., Tsang, A. and Storms, R.K.
Sequence analysis of 68 clones selected at random from an
Aspergillus niger cDNA library
Unpublished (1995)
Contact: Williams BA and Storms RK
Dept of Biology
Concordia University
1455 de Maisonneuve Blvd West, Montreal, Quebec, Canada, H3G 1M8
Tel: (514) 848 3412
Fax: (514) 848 2881
Email: storms@vax2.concordia.ca
(nov01t)

Seq primer: T7 Pro Primer.

FEATURES
Source
1. .268
location/Qualifiers
/organism="Aspergillus niger"
/db_xref="taxon:5061"
/clone_lib="Aspergillus niger, PYES2 (XhoI-EcoRI)"
/lab_host="E. coli"

/note="Vector: PYES2; Site_1: Xho-I; Site_2: EcoRI; cDNA
was synthesized with ZAP kit (Stratagene) using poly(A)+
RNA isolated from Aspergillus niger grown in both complete
and minimal media. Synthesis was primed with oligo(dT)
primer/XhoI-linker. EcoRI adaptors were later ligated to
polished ends. EcoRI-XhoI-digested cDNA was ligated with
EcoRI-XhoI-digested PYES2 (Invitrogen Corp). This vector
permits expression of our library in yeast."

BASE COUNT
ORIGIN
67 a 68 c 69 g 64 t

Query Match 93.3%; Score 14; DB 14; Length 268;
Best Local Similarity 100.0%; Pred. No. 3.6e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTACGGCATGCAC 15
|||||
Db 45 CCTACGGCATGCAC 58

RESULT 8
A1764493/c 325 bp mRNA linear EST 25-JUN-1999
LOCUS UI-R-Y0-abk-a-10-0-UI.s1 UI-R-Y0 Rattus norvegicus cDNA clone
DEFINITION UI-R-Y0-abk-a-10-0-UI 3', mRNA sequence.
A1764493
A1764493.1 GI:5210386

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Norway rat.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
1 (bases 1 to 325)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.iowa.edu
Oligo-dT track not found, Not I site shown in beginning of sequence
is likely internal to the message. cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 95-157, >(CGG
)n#Simple_repeat
Seq primer: M13 Forward
POLYA=NO.

FEATURES
Source
1. .325
location/Qualifiers

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-Y0-abk-a-10-0-UI"
/clone_lib="UI-R-Y0"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-Y0
library is a subtracted library derived from an
individually-tagged normalized whole-eye (minus the lens)
library. The driver for the subtraction consisted of a
pool of all previous libraries (UI-R-A0, UI-R-A1, UI-R-E0,
UI-R-E1, UI-R-C0, and UI-R-C1). The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted

library (UI-R-Y0) was constructed as follows: PCR amplified cDNA inserts from previous library clones from which 3' ESTs had been derived were used as a driver in a hybridization with the normalized whole-eye library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-Y0 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)

 TAG_LIB=UI-R-Y0
 TAG_TISSUE=Eye
 TAG_SEQ="CATTG"

BASE COUNT	37	a	118	c	126	g	43	t	1	others
ORIGIN										
Query Match	93.3%; Score 14; DB 9; Length 325;									
Best Local Similarity	100.0%; Pred. No. 3.9e+03;									
Matches	14;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	2	CCTACGGCATGCAC	15							
DB	174	CCTACGGCATGCAC	161							

```

RESULT 9
BI183951
LOCUS
DEFINITION
BI183951 369 bp mRNA linear EST 10-JUL-2001
UNL-P-FN-cf-e-03-0-UNL.s1 UNL-P-FN Sus scrofa cDNA clone
UNL-P-FN-cf-e-03-0-UNL 3', mRNA sequence.
ACCESSION
BI183951
VERSION
BI183951.1 GI:14658360
KEYWORDS
EST.
SOURCE
pig.
ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE
1 (bases 1 to 369)
Caetano,A.R., Johnson,R.K. and Pomp,D.
Generation and sequence characterization of a normalized cDNA
library from swine ovarian follicles
unpublished (2001)
JOURNAL
Contact: Pomp, D
COMMENT
Department of Animal Science
University of Nebraska, Lincoln
Lincoln, NE 68583-0908, USA
Tel: 402 472 6416
Fax: 402 472 6362
Email: dpomp@unl.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
normalized porcine ovarian follicles library The following
repetitive elements were found in this cDNA sequence: 258-349,
>gll12087|emb1X64127.1|SSPRE S.scrofa DNA for SINE sequence SSPRE
Seq primer: M13 -29
POLYA=Yes.
FEATURES
source
Location/Qualifiers
1..369
/organism="Sus scrofa"
/strain="University of Nebraska, Lincoln Swine Selection
lines"
/db_xref="taxon:9823"
/clone="UNL-P-FN-cf-e-03-0-UNL"
/clone_lib="UNL-P-FN"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UNL-P-FN

```

library is a normalized library representing porcine ovarian follicles, ranging between 2.0 to 10.0 mm in diameter, collected during 7 days of the follicular phases of the pig estrous cycle. This library was derived from the library UNL-P-F2. The tag is a string of 5-6 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996.

```

TAG_LIB=UNL-P-FN
TAG_TISSUE=porcine ovarian follicles
TAG_SEQ=CACACT"

```

BASE COUNT	86 a	100 c	99 g	84 t
ORIGIN				
Query Match				93.3%; Score 14; DB 13; Length 369;
Best Local Similarity				100.0%; Pred. No. 4.1e+03;
Matches	14; Conservative	0;	Mismatches	0; Indels 0; Gaps 0;
QY	1 ACCTACGGCATGCA 14			
Db	264 ACCTACGGCATGCA 277			

RESULT	10
LOCUS	BM728351/c 486 bp mRNA linear EST 01-MAR-2002
DEFINITION	UI-E-EOI-aiv-i-15-0-UI.r1 UI-E-EOI Homo sapiens cDNA clone
ACCESSION	BM728351
VERSION	BM728351
KEYWORDS	GI:19049684
SOURCE	EST.
ORGANISM	human. Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 486)
AUTHORS	Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL MEDLINE COMMENT	Genome Res. 6 (9), 791-806 (1996) 97044477 Contact: Soares, MB Program for Rat Gene Discovery and Mapping University of Iowa 451 Eckstein Medical Research Building Iowa City, IA 52242, USA Tel: 319 335 8250 Fax: 319 335 9565 Email: msocares@blue.weeg.uiowa.edu Tissue Procurement: Dr. Gregg Hageman CDNA Library preparation: Dr. M. Bento Soares, University of Iowa cDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com). Seq primer: MJ3 Reverse.
FEATURES	
source	Location/Qualifiers l..486 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="UI-E-EOI-aiv-i-15-0-UI" /clone_lib="UI-E-EOI" /tissue_type="fetal eye" /dev_stage="fetal" /lab_host="DH10B (Life Technologies) (T1 phage resistant))" /note="Organ: eye; Vector: pTZ19-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-E-EOI is a normalized CDNA library containing the following tissue(s): fetal eye. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand CDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded CDNA was ligated to an EcoR I"

adaptor, digested with Not I, and cloned directionally into pT7m3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGCGTATACC. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI)."

BASE COUNT 110 a 110 c 136 g 127 t 3 others

ORIGIN

Query Match 93.3%; Score 14; DB 14; Length 486;
Best Local Similarity 93.3%; Pred. No. 4.5e+03;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCAC 15
|||
444 ACNTACGGCATGCAC 430

Db

RESULT 11 531 bp mRNA linear EST 17-APR-2002
BQ040153
LOCUS gdl2b08.y1 Moss EST library PPS Physcomitrella patens cDNA clone
DEFINITION PEP_SOURCE_ID:PPS40315 5' similar to TR:Q41050 Q41050 CORE PROTEIN.
; mRNA sequence.
BQ040153
BQ040153.1 GI:19784287
EST.
Physcomitrella patens.
Physcomitrella patens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Bryophyta;
Bryopsida; Funariidae; Funariales; Funariaceae; Physcomitrella.
1 (bases 1 to 531)
Quatrano, R., Bashlades, S., Cove, D., Cuning, A., Knight, C., Clifton
, S., Maria, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Underwood
, K., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
Steptoe, M., Gibbons, M., Harvey, N., Ritter, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Leeds/Wash U Moss EST Project
Unpublished (1999)
Other_ESTs: gdl2b08.x1
Contact: Ralph Quatrano
Leeds/Wash U Moss EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Libraries were constructed by Dr. Stavros Bashlades as part of the
Physcomitrella EST program (PEP) at the Univ. of Leeds (UK) and
Washington Univ. in St. Louis (USA) DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Celia Knight (c.d.knight@leeds.ac.uk)
High quality sequence stop: 421.

FEATURES
source
1..531
location/Qualifiers
/organism="Physcomitrella patens"
/db_xref="taxon:3218"
/clone="PEP_SOURCE_ID:PPS40315"
/clone_lib="Moss EST library PPS"
/dev_stage="protonemata, 7day old untreated"
/lab_host="E.coli DH10b"
/note="Vector: Bluescript SK-; Site_1: XhoI; Site_2:
EcoRI; Library constructed by Stavros Bashlades and
re-arrayed by A. Cuning & Honglin Rong. Construction of
the cDNA library was carried out using Stratagene's 'UnizAP
- cDNA synthesis kit' to ligate cDNA directionally in
UnizAP XR vector arms. The vector is designed containing
the pBluescript sequence as well as the lambda DNA and
cDNA is cloned in the EcoRI and XhoI sites in the
pBluescript sequence. The vector was then packaged using
Gold gigapackaging extracts, propagated in XL-IBLue MRF

cells and amplified. The library was excied by mass excision using Stratagene's Mass excision kit to infect SOLR cells with phagemids and ampicillin resistant transformants selected. Approximately 1,000,000 colonies were grown and recovered by using Qulagen midi prep kit.2 micro grams of plasmid DNA were used to transform DH10b cells by electroporation. Clones corresponding to abundant transcripts were identified by colony hybridization and eliminated from the library, be rearraying. This library is non-directionally cloned."

BASE COUNT 104 a 135 c 145 g 147 t

ORIGIN

Query Match 93.3%; Score 14; DB 14; Length 531;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACCTACGGCATGCA 14
|||
350 ACCTACGGCATGCA 363

Db

RESULT 12 553 bp DNA linear GSS 23-JUN-1999
AQ657745
LOCUS Sheared DNA-27J6.TF Sheared DNA Trypanosoma brucei genomic clone
DEFINITION Sheared DNA-27J6, DNA sequence.
AQ657745
AQ657745.1 GI:5165513
GSS.
Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 553)
El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library
Unpublished (1999)
Other_GSSs: Sheared DNA-27J6.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tldb/mdb/tbdb/>.
Seq primer: M13-Forward
Class: shotgun.

FEATURES
source
1..553
location/Qualifiers
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-27J6"
/clone_lib="Sheared DNA"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Borell, Oxford University
press, 1999)."

BASE COUNT 182 a 149 c 107 g 115 t
ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 553;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCA 14
|||||
Db 325 ACCTACGGCATGCA 338

RESULT 13
AV386289 569 bp mRNA linear EST 27-OCT-1999
LOCUS AV386289 Halocynthia roretzi Fertilized egg Halocynthia roretzi
DEFINITION cDNA clone 009B23_3', mRNA sequence.
ACCESSION AV386289
VERSION AV386289.1 GI:6131346
KEYWORDS EST.
SOURCE Halocynthia roretzi.
ORGANISM Halocynthia roretzi
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
Stolidobranchia; Pyuridae; Halocynthia.
REFERENCE 1 (bases 1 to 569)
AUTHORS Makabe, K.W.
TITLE Halocynthia roretzi EST
JOURNAL Unpublished (1999)
COMMENT Contact: Kazuhiro W. Makabe
Department of Zoology, Graduate School of Science
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4095
Fax: 81-75-705-1113
Email: kwmakabe@ascidian.zool.kyoto-u.ac.jp.

FEATURES
source
1..569
/organism="Halocynthia roretzi"
/db_xref="taxon:7729"
/clone="009B23_3"
/dev_stage="Fertilized egg"
/note="Organ: embryo"

BASE COUNT 157 a 126 c 124 g 162 t
ORIGIN

Query Match 93.3%; Score 14; DB 10; Length 569;
Best Local Similarity 100.0%; Pred. No. 4.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CCTACGGCATGCAC 15
|||||
Db 160 CCTACGGCATGCAC 173

RESULT 14
AQ943516 591 bp DNA linear GSS 27-JAN-2000
LOCUS AQ943516
DEFINITION Sheared DNA-53M15.TR Sheared DNA Trypanosoma brucei genomic clone
ACCESSION AQ943516
VERSION AQ943516.1 GI:6766781
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 591)
AUTHORS El-Sayed, N., Zhao, S., Zhao, H., Gill, S., Suh, E., Malek, J., Fujii, C.,
Gerrard, C., Leech, V., de Jong, P., Ullu, E., Melville, S., Donelson, J.,
Fraser, C. and Adams, M.
TITLE Determination of clone end sequences from Trypanosoma brucei GUTat
10.1 sheared DNA library

JOURNAL Unpublished (1999)
COMMENT Other_GSSs: Sheared DNA-53M15.TR
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
Clones are derived from the Trypanosoma brucei GUTat 10.1 sheared
DNA library constructed at TIGR. Clones will be available for
distribution through ATCC. Sheared DNA end sequences search page:
<http://www.tigr.org/tdb/mdb/tbdb/>.
Seq primer: M13-Reverse
Class: shotgun.

FEATURES
source
1..591
/organism="Trypanosoma brucei"
/strain="TREU927/4 GUTat 10.1"
/db_xref="taxon:5691"
/clone="Sheared DNA-53M15"
/note="Vector: pUC18; Site_1: SmaI; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically
sheared to give a tight size distribution (approx 2 kb).
The v + i method used for the library construction is
described in detail in Smith, H.O. and Venter, J.C.
(Making small insert libraries for whole genome shotgun
sequencing projects. In Genome Sequencing: A Practical
Approach, eds. M. Vaudin and B. Barrell, Oxford University
Press, 1999)."

BASE COUNT 154 a 158 c 142 g 136 t 1 others
ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 591;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ACCTACGGCATGCA 14
|||||
Db 82 ACCTACGGCATGCA 95

RESULT 15
AZ103592 591 bp DNA linear GSS 09-MAY-2000
LOCUS RPCI-23-21H20.TJ RPCI-23 Mus musculus genomic clone RPCI-23-21H20,
DEFINITION DNA sequence.
ACCESSION AZ103592
VERSION AZ103592.1 GI:7756650
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 591)
AUTHORS Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from

BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics ([info@resgen.com](http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html)). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 21 row: H column: 20
Seq primer: SP6
Class: BAC ends.

FEATURES
Location/Qualifiers

Source 1..591
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-21H20"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 127 a 133 c 133 g 198 t
ORIGIN

Query Match 93.3%; Score 14; DB 17; Length 591;
Best Local Similarity 100.0%; Pred. No. 4.9e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTACGGCATGCAC 15
|||||
Db 242 CCTACGGCATGCAC 229

Search completed: June 26, 2003, 04:50:14
Job time : 754.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	6	ARI83492	ARI83492 Sequence
2	15	100.0	43523	9	AC024250	AC024250 Homo sapi
3	15	100.0	50210	9	AC079261	AC079261 Homo sapi
4	15	100.0	59465	8	AY007367	AY007367 Lycopersi
5	15	100.0	91240	2	AC106140	AC106140 Rattus no
6	15	100.0	100685	8	AC023754	AC023754 Arabidops
7	15	100.0	101282	2	AC121474	AC121474 Rattus no
8	15	100.0	101290	9	AC009240	AC009240 Homo sapi
9	15	100.0	108610	2	AC105725	AC105725 Rattus no
10	15	100.0	121212	9	HS135L22	AL031767 Human DNA
11	15	100.0	127400	2	AC131121	AC131121 Mus muscu
12	15	100.0	129547	8	AC006434	AC006434 Genomic s
13	15	100.0	145520	2	AC016036	AC016036 Homo sapi
14	15	100.0	154084	2	AC022311	AC022311 Homo sapi
15	15	100.0	158861	2	AC130009	AC130009 Rattus no
16	15	100.0	165733	2	AC099146	AC099146 Rattus no
17	15	100.0	170717	2	AC128590	AC128590 Rattus no
18	15	100.0	175518	2	AC018846	AC018846 Homo sapi
19	15	100.0	180613	2	AC115064	AC115064 Mus muscu
20	15	100.0	182604	2	AC127446	AC127446 Rattus no
21	15	100.0	182991	2	AC101658	AC101658 Mus muscu
22	15	100.0	205202	2	AC108796	AC108796 Mus muscu
23	15	100.0	207473	2	AC117630	AC117630 Mus muscu
24	15	100.0	221372	2	AC124474	AC124474 Mus muscu
25	15	100.0	227010	2	AC113490	AC113490 Mus muscu
26	14	93.3	127	9	HUMRB	M27258 Human mutan
27	14	93.3	258	9	HSALJ9519	AJ009519 Homo sapi
28	14	93.3	373	8	AY038113	AY038113 Montanoa
29	14	93.3	415	8	AY038114	AY038114 Montanoa
30	14	93.3	415	8	AY038115	AY038115 Montanoa
31	14	93.3	652	11	PM7H5B	AL685230 Penicillii
32	14	93.3	735	11	PM3B9G	AL684904 Penicillii
33	14	93.3	2570	8	AF422179	AF422179 Arabidops
34	14	93.3	2898	1	L1U95056	U95056 Leptospiira
35	14	93.3	3456	3	AY051923	AY051923 Drosophil
36	14	93.3	9908	1	AE013309	AE013309 Methanosa
37	14	93.3	23819	2	AC018173	AC018173 Drosophil
38	14	93.3	36138	1	MLU15187	U15187 Mycobacteri
39	14	93.3	38109	1	MLCB5	Z95151 Mycobacteri
40	14	93.3	42936	3	U53333	U53333 Caenorhabdi
41	14	93.3	60328	9	AL500526	AL500526 Human DNA
42	14	93.3	63178	3	AC004333	AC004333 Drosophil
43	14	93.3	64631	2	AC118478	AC118478 Mus muscu
44	14	93.3	64767	9	U69730	U69730 Homo sapien
45	14	93.3	71320	2	AC095414	AC095414 Rattus no

ALIGNMENTS

RESULT 1
ARI83492
LOCUS ARI83492 15 bp DNA
DEFINITION Sequence 36 from patent US 6342220.
ACCESSION ARI83492
VERSION ARI83492.1 GI:20227461
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Rendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 36 29-JAN-2002;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

```

source              1..15
BASE COUNT          7 a      4 c      2 g      2 t
ORIGIN
Query Match        100.0%; Score 15; DB 6; Length 15;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGCCATACATGAAC 15
Db      1 AGCCATACATGAAC 15

RESULT 2
AC024250/c
LOCUS      AC024250      43523 bp      DNA      linear      PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-684N2 from Y, complete sequence.
ACCESSION  AC024250
VERSION     AC024250.6  GI:11120963
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 43523)
AUTHORS     Sulston,J.E. and Waterston,R.
TITLE       Toward a complete human genome sequence
JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     99063792
PUBMED      9847074

REFERENCE   2 (bases 1 to 43523)
AUTHORS     Kyung,K., Drone,K. and Le,T.
TITLE       The sequence of Homo sapiens BAC clone RP11-684N2
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 43523)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (28-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            4 (bases 1 to 43523)
            Waterston,R.
            Direct Submission
            Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            5 (bases 1 to 43523)
            Waterston,R.
            Direct Submission
            Submitted (09-MAY-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
            On Nov 8, 2000 this sequence version replaced gi:9958320.

COMMENT     ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watsn.wustl.edu
            ----- Summary Statistics
            Center project name: H_NH0684N02
            -----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
```

```

MAPING INFORMATION:
The position of this clone was established as part of a
collaboration between the Human Chromosome Y Mapping Project
(Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen,
and David C. Page at the Whitehead Institute for Biomedical
Research, Cambridge MA) and the Washington University Genome
Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
Iateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at the Roswell Park Cancer Institute
(http://bacpac.med.buffalo.edu)
VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-943F15; the clone sequenced
to the right is RP11-489O13. Actual start of this clone is at base
position 17386 of RP11-943F15; actual end is at base position 71706
of RP11-489O13.

The sequence fidelity of RP11-684N2 between bases 832 and 43523 can
not be guaranteed due to a tandem repeat. This region contains low
quality, single stranded areas, and base mismatches. Digest
information corresponds to the current assembly.

FEATURES
source
1..43523
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-684N2"
/clone_1lb="RPCI-11"
1..43507
repeat_region
/rpt_family="ERV1"
BASE COUNT      9246 a      8789 c     10023 g     15465 t
ORIGIN
Query Match        100.0%; Score 15; DB 9; Length 43523;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGCCATACATGAAC 15
Db      17764 AGCCATACATGAAC 17750

RESULT 3
AC079261/c
LOCUS      AC079261      50210 bp      DNA      linear      PRI 09-MAY-2001
DEFINITION Homo sapiens BAC clone RP11-1325K3 from Y, complete sequence.
ACCESSION  AC079261
VERSION     AC079261.2  GI:11024958
KEYWORDS    HTG.
SOURCE      Homo sapiens.
ORGANISM    Homo sapiens.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1 (bases 1 to 50210)
AUTHORS     Sulston,J.E. and Waterston,R.
TITLE       Toward a complete human genome sequence
JOURNAL     Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE     9847074
PUBMED      9847074

REFERENCE   2 (bases 1 to 50210)
AUTHORS     Ozersky,P., Abbott,A., Maupin,R. and Ames,M.
TITLE       The sequence of Homo sapiens BAC clone RP11-1325K3
JOURNAL     Unpublished
REFERENCE   3 (bases 1 to 50210)
```

AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (25-AUG-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 50210)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (26-OCT-2000) Genome Sequencing Center, Washington,
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 50210)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 26, 2000 this sequence version replaced gi:9910108.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH1325K03

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)
VECTOR: PBACE3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-1285C3; the clone sequenced to the right is RP11-943F15. Actual start of this clone is at base position 1 of RP11-1325K3; actual end is at base position 50210 of RP11-1325K3.

The entire clone of H_NH1325K03 from 1 to 50210, is a 125 bp unit tandem repeat. There are low quality areas, single clone areas, single chemistry areas and high quality mismatches throughout this clone. Digest cannot confirm any data within this clone due to the size of the tandem and its repetitive nature. Due to the tandem the exact overlap cannot be determined, but RP11-1325K3 and RP11-943F15 are thought to overlap. The two clones are going to be butt joined to give the best available representation of the region.

FEATURES

Location/Qualifiers

source 1..50210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-1325K3"
/clone_1lb="RPCI-11"
12..50203
repeat_region /rpt_family="ERV1"
BASE COUNT 10786 a 9972 c 11695 g 17757 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 50210;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches .0; Indels 0; Gaps 0;
QY 1 ACCCATACATGAAC 15
|||||
Db 24811 AGCCATACATGAAC 24797
RESULT 4
AY007367 59465 bp DNA linear PLN 01-SEP-2001
LOCUS
DEFINITION Lycopersicon esculentum tospovirus resistance protein C (Sw5-c),
tospovirus resistance protein D (Sw5-d), and tospovirus resistance
protein E (Sw5-e) genes, complete cds.
ACCESSION AY007367
VERSION
KEYWORDS
SOURCE AY007367.1 GI:15418711
ORGANISM Lycopersicon esculentum.
Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
Lycopersicon.
1 (bases 1 to 59465)
Folkertsma,R.T., Spassova,M.I., Prins,M., Stevens,M.R., Hille,J.
and Goldbach,R.W.
Construction of a bacterial artificial chromosome (BAC) library of
Lycopersicon esculentum cv. Stevens and its application to
physically map the Sw-5 locus
Mol. Breed. 5, 197-207 (1999)
2 (bases 1 to 59465)
Prins,T.W., Spassova,M.I., Prins,M., Klein lankhorst,R.,
Folkertsma,R.T., Hille,J. and Goldbach,R.W.
The Sw-5 locus confers resistance to Tospoviruses
Unpublished
JOURNAL 3 (bases 1 to 59465)
REFERENCE Prins,T.W., Spassova,M.I., Prins,M., Klein lankhorst,R.,
Folkertsma,R.T., Hille,J. and Goldbach,R.W.
TITLE Direct Submission
JOURNAL Submitted (22-AUG-2000) Department of Virology, Wageningen
University, Binnenhaven 11, Wageningen 6709 PD, The Netherlands
FEATURES
source 1..59465
/organism="Lycopersicon esculentum"
/cultivar="Stevens"
/db_xref="taxon:4081"
/chromosome="9"
/map="tightly linked to RFLP marker CT220"
<23512..>27327
/gene="Sw5-c"
<23512..>27327
/gene="Sw5-c"
/product="tospovirus resistance protein C"
23512..27327
/note="L2-NBS-LRR class resistance protein"
/codon_start=1
/product="tospovirus resistance protein C"
/protein_id="AG31015.1"
/db_xref="GI:15418712"

Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sulton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Unpublished
Direct Submission
2 (bases 1 to 91240)
Worley,K.C.

Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 91240)
Worley,K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:18138661.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-helpebcm.tmc.edu
----- Project Information
Center project name: GJGR
Center clone name: CH230-230F16
----- Summary Statistics
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap: version 0.990329
Consensus quality: 36346 bases at least Q40
Consensus quality: 39251 bases at least Q30
Consensus quality: 42076 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 54 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1264: contig of 1264 bp in length
* 1265 1364: gap of unknown length
* 1365 2650: contig of 1286 bp in length
* 2651 2750: gap of unknown length
* 2751 3941: contig of 1191 bp in length
* 3942 4041: gap of unknown length
* 4042 5286: contig of 1245 bp in length
* 5287 5386: gap of unknown length
* 5387 6670: contig of 1284 bp in length
* 6671 6770: gap of unknown length
* 6771 7909: contig of 1139 bp in length
* 7910 8009: gap of unknown length
* 8010 9162: contig of 1153 bp in length
* 9163 9262: gap of unknown length
* 9263 10570: contig of 1308 bp in length
* 10571 10670: gap of unknown length
* 10671 11724: contig of 1054 bp in length
* 11725 11824: gap of unknown length
* 11825 12998: contig of 1174 bp in length
* 12999 13098: gap of unknown length

* 13099 14265: contig of 1167 bp in length
* 14266 14365: gap of unknown length
* 14366 15589: contig of 1224 bp in length
* 15590 15689: gap of unknown length
* 15690 16720: contig of 1031 bp in length
* 16721 16820: gap of unknown length
* 16821 17903: contig of 1083 bp in length
* 17904 18003: gap of unknown length
* 18004 19222: contig of 1219 bp in length
* 19223 19322: gap of unknown length
* 19323 20454: contig of 1132 bp in length
* 20455 20554: gap of unknown length
* 20555 21851: contig of 1297 bp in length
* 21852 21951: gap of unknown length
* 21952 23915: contig of 1964 bp in length
* 23916 24015: gap of unknown length
* 24016 25805: contig of 1790 bp in length
* 25806 25905: gap of unknown length
* 25906 27163: contig of 1258 bp in length
* 27164 27263: gap of unknown length
* 27264 28571: contig of 1308 bp in length
* 28572 28671: gap of unknown length
* 28672 30050: contig of 1379 bp in length
* 30051 30150: gap of unknown length
* 30151 31771: contig of 1621 bp in length
* 31772 31871: gap of unknown length
* 31872 33132: contig of 1261 bp in length
* 33133 33232: gap of unknown length
* 33233 34592: contig of 1360 bp in length
* 34593 34692: gap of unknown length
* 34693 35740: contig of 1048 bp in length
* 35741 35840: gap of unknown length
* 35841 37229: contig of 1389 bp in length
* 37230 37329: gap of unknown length
* 37330 38840: contig of 1511 bp in length
* 38841 38940: gap of unknown length
* 38941 40795: contig of 1855 bp in length
* 40796 40895: gap of unknown length
* 40896 43165: contig of 2270 bp in length
* 43166 43265: gap of unknown length
* 43266 44613: contig of 1348 bp in length
* 44614 44713: gap of unknown length
* 44714 45767: contig of 1054 bp in length
* 45768 45867: gap of unknown length
* 45868 47485: contig of 1618 bp in length
* 47486 47585: gap of unknown length
* 47586 49220: contig of 1635 bp in length
* 49221 49320: gap of unknown length
* 49321 50434: contig of 1114 bp in length
* 50435 50534: gap of unknown length
* 50535 52376: contig of 1842 bp in length
* 52377 52476: gap of unknown length
* 52477 54365: contig of 1889 bp in length
* 54366 54465: gap of unknown length
* 54466 55482: contig of 1017 bp in length
* 55483 55582: gap of unknown length
* 55583 57326: contig of 1744 bp in length
* 57327 57426: gap of unknown length
* 57427 58482: contig of 1056 bp in length
* 58483 58582: gap of unknown length
* 58583 60390: contig of 1808 bp in length
* 60391 60490: gap of unknown length
* 60491 62642: contig of 2152 bp in length
* 62643 62742: gap of unknown length
* 62743 64159: contig of 1417 bp in length
* 64160 64259: gap of unknown length
* 64260 65435: contig of 1176 bp in length
* 65436 65535: gap of unknown length
* 65536 68178: contig of 2643 bp in length
* 68179 68278: gap of unknown length
* 68279 71375: contig of 3097 bp in length
* 71376 71475: gap of unknown length
* 71476 73359: contig of 1884 bp in length

* 73360	73459: gap of unknown length
* 73460	75813: contig of 2354 bp in length
* 75814	75913: gap of unknown length
* 75914	77801: contig of 1888 bp in length
* 77802	77901: gap of unknown length
* 77902	79798: contig of 1897 bp in length
* 79799	79898: gap of unknown length
* 79899	82762: contig of 2864 bp in length
* 82763	82862: gap of unknown length
* 82863	86400: contig of 3538 bp in length
* 86401	86500: gap of unknown length
* 86501	88336: contig of 1836 bp in length
* 88337	88436: gap of unknown length
* 88437	91240: contig of 2804 bp in length.

Query Match	100.0%;	Score 15;	DB 2;	Length 91240;
Best Local Similarity	100.0%;	Pred. No. 2e+02;		
Matches 15;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 AGCCATAACATGAAC 15
Db 88261 AGCCATAACATGAAC 88247

RESULT 6
AC023754
LOCUS AC023754 100685 bp DNA linear PLN 13-SEP-2000
DEFINITION Arabidopsis thaliana chromosome I BAC F1B16 genomic sequence,
AC023754 complete sequence.
AC023754
VERSION AC023754.3 GI:7212003
KEYWORDS HTG.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE
AUTHORS
1 (bases 1 to 100685)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Buehler,E., Chao,Q., Chin,C., Chio,J., Choi,E., Gonzalez,A.,
Howng,B., Johnson-Hopson,C., Khan,S., Kim,C., Koo,T., Lee,J.M.,
Lenz,C., Liu,A., Liu,S., Mukharsky,N., Pham,P., Sakano,H.,
Shinn,P., Toriumi,M., Vaysberg,M., Yu,G., Ecker,J., Theologis,A.
and Davis,R.W.
Unpublished

JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 100685)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chio,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission

TITLE
JOURNAL
Submitted (17-FEB-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
3 (bases 1 to 100685)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Bei,Q., Buehler,E.,
Chin,C., Chio,J., Choi,E., Dunn,P., Gonzalez,A., Howng,B., Kim,C.,
Koo,T., Lee,J.M., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S.,
Mukharsky,N., Pham,P., Sakano,H., Schwartz,J., Shinn,P.,
Thaveri,A., Toriumi,M., Vaysberg,M., Walker,M., Yu,G., Ecker,J.,
Theologis,A. and Davis,R.W.
Direct Submission

TITLE
JOURNAL
Submitted (09-MAR-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
4 (bases 1 to 100685)
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,
Altafi,H., Nguyen,M., Lam,B., Southwick,A., Ecker,J., Theologis,A.

TITLE
JOURNAL
Submitted (13-SEP-2000) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT
On Mar 9, 2000 this sequence version replaced gi:7143418.
Bases 1-32,040 of IGF clone F1B16 overlap with bases 97,508-129,547
of IGF clone F10A5, gb|AC006434.
Genes with similarity to proteins in the databases are named
'putative', '-like' or 'similar to'. Genes that have EST
similarity but no significant protein similarity are described as
'unknown proteins'. Genes that are annotated based only on gene
prediction software are described as 'hypothetical proteins'. The
gene prediction programs used to predict genes include: Grail
(Informatics Group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge,
<http://genes.mit.edu/GENSCAN/>), FEXA (Victor Solovyev,
<http://genomic.sanger.ac.uk/gf/gf.shtml>), and NetPlantGene (S.M.
Hebsgaard, et al., CBS, Technical University of
Denmark, <http://www.cbs.dtu.dk/NetPlantGene.html>).

FEATURES
source

misc_feature

gene
CDS

gene
CDS

gene
CDS

1. 100685
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="I"
/clone="F1B16"
1. 32040
/note="overlap with bases 97,508-129,547 of IGF clone
F10A5, gb|AC006434. See GenBank record for BAC clone
F10A5 for annotation in this region."
complement(32986..33822)
/gene="F1B16.1"
complement(32986..33822)
/gene="F1B16.1"
/note="Unknown protein; unknown protein"
/codon_start=1
/protein_id="AAG13069.1"
/db_xref="GI:10120444"
/translation="MALPHVSSPSSSTLSPSFKRPFHSLRSLSPYVDRRCRLN
SSFRCSSESEKHNNAKSPKSDIVELPLPLPLVLPFGAILPLQIFEFRTIMHT
LLQSDLRFGVVYSDSVSGSAEVEGVEVKKHERLVDRFLLVCKGQERFRVTVRT
KPYLGEVTWLEDPRSGEENDSLANEVEVLMKEVIRLSNRLNGKAEKEVQDLRRNQF
PTPFSEFVGSTFEGAPREQALLLEEDTAARLKRERTLRNTLNYLTAASAVKDYFPS
SS"
43756..46560
/gene="F1B16.2"
join(43756..44347,44507..44562,44700..44966,45493..45755,
46200..46560)
/gene="F1B16.2"
/note="Similar to cytokinin oxidase"
/codon_start=1
/protein_id="AAG13068.1"
/db_xref="GI:10120443"
/translation="MTSSFLLLTFAICKLIIVAGLNVGPSELLRIGAVDGHFTVHP
SDLASVSSDFGMLKSPEEPLAVLHPSSAEDVARLVRTAYGSATAPFVSARGHGSING
QAAAGRNGVVVENMNGVGTGPKPLVRPDENVYDVWGGELMDVLKKTLEHGLAPKSWT
DYLALTVGGTLSNAGISGQAFHHGPQISNVLELDVVTGQGITTRARISLEPAPQVR
WIRLVYSSFKVFTEDEQYILSMHGOLKFDYVEGVIVDEGLVNNRRSSFSPPRPVKI
SSVSSNGSVLYCLEITTKNYHDSDEIVDQVEILMKLNFIPTSVFTTDLQYVDFDR
VHKAEILKRSKLMVEVPHWLNLFVPKSRISDRDKGVFKILGNKTSGPILILYPMNKD
KWDERSAATVPDEVEFYVALRLSALTDEETQKLEYLKDQNRILFEQAKINVKO
YLPNHATQEEVNAHFQDKWDRFRSLKAIEFDRHILATGQIRIFQNPSTSLFPPSSSSSS
AASW"
complement(46848..48185)
/gene="F1B16.3"
complement(join(46848..46943,47040..47128,47474..47537,
47773..47875,47963..48066,48156..48185))
/gene="F1B16.3"
/codon_start=1
/product="Putative ubiquitin-conjugating enzyme"
/protein_id="AAG13066.1"
/db_xref="GI:10120441"

gene
CDS
52291. .53501
/gene="F1B16.4"
join(52291. .52584,52917. .53501)
/gene="F1B16.4"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAG13063.1"
/db_xref="GI:10120438"
/translation="MEDFRVRHCESSLRCGLLDLSRYAKAVQCLVEEVIDIGREVELC
NNILINOLFPGRRRPGFALSSIEKSELSSGFMSLPENHEIHIKITKLSDLQVEER
FEQYCNQLEQYISSFEEIAGEGSSKVYTGIALQAMTRHFGSLEAIIISQNSVRRFI
ISHQDPKTISSGLQSLDFDGNITSSSLQRLGLVQGPORHAWKPIRGLEPESVALR
AWLFQHLRPSVFLLLKCNNTSDESKLCKMKNNISLIAVMKTVIRMKQRSWCWRLKQD
FRRTKVTPLNLLLF"
complement(54491. .56307)
/gene="F1B16.5"
complement(join(54491. .55074,55197. .55351,55498. .55769,
55872. .56014,56253. .56307))
/gene="F1B16.5"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAG13070.1"
/db_xref="GI:10120445"
/translation="MGTDGFVVSICINSYDSCRGPLLMEFLRLRGVADVYVITNQ
KPLEDEDEVVYSLHEKMLDRGVQVISAKGQKAVDTSLKADLIVLTAVAGKWLDAVLKE
NVKVLPKILMWIHEMRGHYFNADVLKHLPEVAGAMIDSHTAGYWKNRTPQARLGIKM
PKTYVHLGNSKELMEVAEDSVAKRLREHVRRESLGRNEDLFGIINSVSRGGODL
FLRAFHESLERIEKKLQVPTMAVAVGSDMSKQTELELRFNVRREKLENEVHFVN
KTLTVAPYIAIDVLVQNSQARGECFRITTEAMAFKLPVLGTAAGTMEIVVNGTTG
LHSAGKEGVIPLAKNIVKLATQVELRLRMGKNGYERVKEMFLEHMHSHRIASVLYEV
LOHAKARALR"
59918. .61713
/gene="F1B16.6"
join(59918. .60655,60734. .61110,61188. .61248,61315. .61713)
/gene="F1B16.6"
/note="Similar to homeodomain proteins"
/codon_start=1
/protein_id="AAG13065.1"
/db_xref="GI:10120440"
/translation="MAVYYPNSVGMOSLYQESTYLNEQOQQQQOASSSSAASFSEIVS
GDVRNNEVFIPPTSDVAVNGNVTSSNDLSFHGGSLSLGNQIOSAVSVSPQYHY
QNLNQLSYNNLNPSTMSDENGKSLSVHQHSDQILPSSVYNNNGNGVGYNNRYE
TSGFVSSVLSRXYLKPTQQLDEVSVRKDLKLGKMKMKNDKGODFHNGSDNITEDD
KSQSELSPSERQELQSKSLITMVEVDKRYNQYHHQMEALASSFEMVTGLGAAP
YTSVALNIRISRHFRLDAIKEQIOIVIRGLGERETSDQGERIPRLRYIDQRLROOR
ALHQQLGMVRPAWRPQRPQDPENSVISLRWLFEHFLHPYPKSEKIMLSKQTLGSKNQ
VANWFINARVRLWKPMEIEMYKEEFGESALLSNSNDTKMKQETSQLKHEDESSSQO
QNGNNNNNNIPTYSDAEQNLVFAFPKPRDATTGDYDSLNNYHGFGIDDYNNRYVGLNQ
QDGRYSNPQHLHDFV"
complement(62425. .64698)
/gene="F1B16.7"
complement(join(62425. .63429,63528. .63559,63863. .64074,
64640. .64698))
/gene="F1B16.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="AAG13067.1"
/db_xref="GI:10120442"
/translation="MPFRIYSCFDEPRNLSLFWRRVADEIEGSPYSPYVASRGISM
DKMSLGSERGLSEGGTPPDCHLGT PASQKATPEMSTNSMVPSSGSSSLASDLIEV
KNLYDSPGIVPSAOSKPLPFSPLPNPVSDDLSSAHTRLPDKTSPSRARSPGHOLF
RQYSDSQILGLKSPINNYSISEGRSSFVLSFSCSDNPFATGSQYASSGEGWSMNTFSELY
AYQSDRKSFDSEHLGSGRRKLGGSSRSFSPSVVDQVCGACSKLITERSIATFE
LPIAAVLACGHVYHAECLFETMTDIEKYDPACPICTIGEKRYAKITRKALKAEAEAKA
KOYRCRKNRVVDSYGESECEDEVYQKMGKREGKALKLEASCSSSSSKSFLKWHFAS
ISSKWNKPSKSDSALKKGFWSRHRNRRSSSIEVKIHTLNIQS"
complement(69063. .69566)
/gene="F1B16.8"
complement(69063. .69566)
CDS

```

/gene="Fib16.8"
/note="similar to DNA-binding protein"
/codon_start=1
/protein_id="AAG13064.1"
/db_xref="GI:10120439"
/translation="MGSSTGNCSSVSTTGLANGSESDLRORDLIDERKRRKQSNR
ESARSRMRKQHLLDILTAVTHLKRKNQAIVAGIAVTQTHTYITAEANDILRAQYLE

Query Match          100.0%; Score 15; DB 8; Length 100685;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAAC 15
        |||
Db       16468 AGCCATAACATGAAC 16482

RESULT 7
AC121474 LOCUS
DEFINITION AC121474 Rattus norvegicus clone CH230-88M5, *** SEQUENCING IN PROGRESS ***,
62 unordered pieces.
AC121474 AC121474.2 GI:21909158
VERSION HTG; HTGS_PHASE1.
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 101282)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhaq,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleaveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Hollway,C., Hollins,B.,
Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokwenkwu,S., Oguh,M., Okwuonu,G.,
Orangunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Taborek,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlarczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
```

REFERENCE 2 (bases 1 to 101282)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 101282)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT On Jul 19, 2002 this sequence version replaced gi:20976426.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYV
Center clone name: CH230-88M5
----- Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 53833 bases at least Q40
Consensus quality: 58211 bases at least Q30
Consensus quality: 60723 bases at least Q20.

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 62 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1096: contig of 1095 bp in length
* 1096 1195: gap of unknown length
* 1196 2364: contig of 1169 bp in length
* 2365 2464: gap of unknown length
* 2465 3475: contig of 1011 bp in length
* 3476 3575: gap of unknown length
* 3576 4593: contig of 1018 bp in length
* 4594 4693: gap of unknown length
* 4694 5854: contig of 1161 bp in length
* 5855 5954: gap of unknown length
* 5955 6984: contig of 1030 bp in length
* 6985 7084: gap of unknown length
* 7085 8190: contig of 1106 bp in length
* 8191 8290: gap of unknown length
* 8291 9708: contig of 1418 bp in length
* 9709 9808: gap of unknown length
* 9809 11270: contig of 1462 bp in length
* 11271 11370: gap of unknown length
* 11371 12681: contig of 1311 bp in length
* 12682 12781: gap of unknown length
* 12782 13856: contig of 1075 bp in length
* 13857 13956: gap of unknown length
* 13957 15348: contig of 1392 bp in length
* 15349 15448: gap of unknown length
* 15449 16775: contig of 1327 bp in length
* 16776 16875: gap of unknown length
* 16876 18345: contig of 1470 bp in length
* 18346 18445: gap of unknown length
* 18446 19882: contig of 1437 bp in length
* 19883 19982: gap of unknown length
* 19983 21546: contig of 1564 bp in length
* 21547 21646: gap of unknown length
* 21647 23246: contig of 1600 bp in length
* 23247 23346: gap of unknown length

* 23347 24676: contig of 1330 bp in length
* 24677 24776: gap of unknown length
* 24777 26040: contig of 1264 bp in length
* 26041 26140: gap of unknown length
* 26141 27293: contig of 1153 bp in length
* 27294 27393: gap of unknown length
* 27394 28872: contig of 1479 bp in length
* 28873 28972: gap of unknown length
* 28973 30187: contig of 1215 bp in length
* 30188 30287: gap of unknown length
* 30288 31484: contig of 1197 bp in length
* 31485 31584: gap of unknown length
* 31585 32606: contig of 1022 bp in length
* 32607 32706: gap of unknown length
* 32707 33988: contig of 1282 bp in length
* 33989 34088: gap of unknown length
* 34089 35495: contig of 1407 bp in length
* 35496 35595: gap of unknown length
* 35596 37634: contig of 2039 bp in length
* 37635 37734: gap of unknown length
* 37735 39181: contig of 1447 bp in length
* 39182 39281: gap of unknown length
* 39282 40774: contig of 1493 bp in length
* 40775 40874: gap of unknown length
* 40875 42029: contig of 1155 bp in length
* 42030 42129: gap of unknown length
* 42130 43727: contig of 1598 bp in length
* 43728 43827: gap of unknown length
* 43828 45082: contig of 1255 bp in length
* 45083 45182: gap of unknown length
* 45183 46312: contig of 1130 bp in length
* 46313 46412: gap of unknown length
* 46413 47654: contig of 1242 bp in length
* 47655 47754: gap of unknown length
* 47755 48861: contig of 1107 bp in length
* 48862 48961: gap of unknown length
* 48962 50268: contig of 1307 bp in length
* 50269 50368: gap of unknown length
* 50369 51686: contig of 1318 bp in length
* 51687 51786: gap of unknown length
* 51787 52914: contig of 1128 bp in length
* 52915 53014: gap of unknown length
* 53015 54429: contig of 1415 bp in length
* 54430 54529: gap of unknown length
* 54530 55937: contig of 1408 bp in length
* 55938 56037: gap of unknown length
* 56038 57649: contig of 1612 bp in length
* 57650 57749: gap of unknown length
* 57750 59809: gap of unknown length
* 59810 59909: contig of 2060 bp in length
* 59910 61896: contig of 1987 bp in length
* 61897 61996: gap of unknown length
* 61997 63329: contig of 1333 bp in length
* 63330 63429: gap of unknown length
* 63430 64748: contig of 1319 bp in length
* 64749 64848: gap of unknown length
* 64849 66048: contig of 1200 bp in length
* 66049 66148: gap of unknown length
* 66149 67773: contig of 1625 bp in length
* 67774 67873: gap of unknown length
* 67874 69469: contig of 1596 bp in length
* 69470 69569: gap of unknown length
* 69570 71241: contig of 1672 bp in length
* 71242 71341: gap of unknown length
* 71342 73081: contig of 1740 bp in length
* 73082 73181: gap of unknown length
* 73182 75302: contig of 2121 bp in length
* 75303 75402: gap of unknown length
* 75403 77624: contig of 2222 bp in length
* 77625 77724: gap of unknown length
* 77725 79324: contig of 1600 bp in length
* 79325 79424: gap of unknown length
* 79425 80888: contig of 1464 bp in length

Query Match 100.0%; Score 15; DB 2; Length 101282;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
|||||

Db 86816 AGCCATACATGAAC 86830

RESULT 8
AC009240/c 101290 bp DNA linear PRI 07-OCT-2000
LOCUS Homo sapiens BAC clone RP11-489013 from Y, complete sequence.
DEFINITION AC009240
ACCESSION AC009240
VERSION AC009240.6 GI:10140836
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 101290)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074

REFERENCE 2 (bases 1 to 101290)
AUTHORS Scott,K., Holmes,A., Maupin,R. and Hodges,J.
TITLE The sequence of Homo sapiens BAC clone RP11-489013
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 101290)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (06-AUG-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 4 (bases 1 to 101290)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

REFERENCE 5 (bases 1 to 101290)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 15, 2000 this sequence version replaced gi:9838226.
COMMENT ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
Center project name: H_NH0489013

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The position of this clone was established as part of a

collaboration between the Human Chromosome Y Mapping Project (Tomoko Kawaguchi, Helen Skaletsky, Laura G. Brown, Steve Rozen, and David C. Page at the Whitehead Institute for Biomedical Research, Cambridge MA) and the Washington University Genome Sequencing Center, St. Louis MO.

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., rateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu)

VECTOR: pBACe3.6
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-684N2, the clone sequenced to the right is RP11-329C15, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-489013; actual end is at base position 73076 of RP11-329C15.

The sequence H_NH0489013 from base position 1 to 32763 contains a tandem repeat which continues into RP11-684N2. The sequence fidelity of the tandem repeat cannot be guaranteed. Digest data was used to make the assembly however, it still suggests there is missing data.

FEATURES
source 1. 101290
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="Y"
/map="Y"
/clone="RP11-489013"
/clone_lib="RPCI-11"
18. .32348
/rpt_family="ERV1"
32385. .32820
/rpt_family="L1"
32810. .33021
/rpt_family="L1"
33046. .34646
/rpt_family="L1"
34914. .35140
/rpt_family="Alu"
35144. .36404
/rpt_family="L1"
36408. .37075
/rpt_family="L1"
37076. .37374
/rpt_family="Alu"
37375. .38180
/rpt_family="L1"
38191. .38328
/rpt_family="L1"
38564. .38835
/rpt_family="L1"
39841. .39894
/rpt_family="Alu"
39920. .40231
/rpt_family="Alu"
40388. .40988
/rpt_family="L1"
40990. .41925
/rpt_family="L1"
41981. .42074
/rpt_family="L1"
42075. .42369
/rpt_family="Alu"
42370. .42630
/rpt_family="L1"
43200. .43400
/rpt_family="L1"

repeat_region 43485. .43853
/rpt_family="L1"
repeat_region 43879. .44034
/rpt_family="L1"
repeat_region 44421. .44873
/rpt_family="MaLR"
repeat_region 44876. .45175
/rpt_family="L1"
repeat_region 45939. .46149
/rpt_family="L1"
repeat_region 46241. .46473
/rpt_family="L1"
repeat_region 46474. .46769
/rpt_family="Alu"
repeat_region 46770. .46945
/rpt_family="L1"
repeat_region 47026. .47143
/rpt_family="L1"
repeat_region 47158. .47345
/rpt_family="Alu"
repeat_region 47891. .48318
/rpt_family="L1"
repeat_region 48332. .48394
/rpt_family="L1"
repeat_region 48773. .49076
/rpt_family="Alu"
repeat_region 55508. .55666
/rpt_family="Alu"
repeat_region 55675. .56009
/rpt_family="L1"
repeat_region 56426. .56710
/rpt_family="Alu"
repeat_region 56753. .57206
/rpt_family="L1"
repeat_region 59131. .59503
/rpt_family="MaLR"
repeat_region 59810. .59935
/rpt_family="Alu"
repeat_region 59949. .59990
/rpt_family="ERVL"
repeat_region 60011. .60189
/rpt_family="Alu"
repeat_region 60294. .60498
/rpt_family="ERV1"
repeat_region 60502. .61724
/rpt_family="L1"
repeat_region 61725. .63600
/rpt_family="L1"
repeat_region 63601. .67996
/rpt_family="L1"
repeat_region 68032. .68464
/rpt_family="L1"
repeat_region 68468. .69183
/rpt_family="L1"
repeat_region 69239. .69529
/rpt_family="Alu"
repeat_region 69737. .70017
/rpt_family="L1"
repeat_region 70050. .70099
/rpt_family="L1"
repeat_region 71896. .73699
/rpt_family="L1"
repeat_region 73726. .73887
/rpt_family="L1"
repeat_region 73998. .74248
/rpt_family="L1"

Query Match 100.0%; Score 15; DB 9; length 101290;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
|||||

Db 12684 AGCCATACATGAAC 12670

RESULT 9
AC105725
LOCUS
DEFINITION AC105725 108610 bp DNA linear HTG 13-JUL-2002
Rattus norvegicus clone CH230-64F5, *** SEQUENCING IN PROGRESS ***,
56 unordered pieces.
AC105725
AC105725.2 GI:21735171
HTG, HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE
AUTHORS
1 (bases 1 to 108610)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Farnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homsli,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., Mcleod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoostari,N., Sisson,I.,
Sodergren,E., Sonaiké,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 108610)
Worley,K.C.
Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 108610)
Worley,K.C.
Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 12, 2002 this sequence version replaced gi:18092948.
----- Genome Center

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GNOM
Center clone name: CH230-64E5
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 67703 bases at least Q40
Consensus quality: 72041 bases at least Q30
Consensus quality: 75764 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 56 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1010: contig of 1010 bp in length
1011 1110: gap of unknown length
1111 2528: contig of 1418 bp in length
2529 2628: gap of unknown length
2629 4257: contig of 1629 bp in length
4258 4357: gap of unknown length
4358 5571: contig of 1214 bp in length
5572 5671: gap of unknown length
5672 6914: contig of 1243 bp in length
6915 7014: gap of unknown length
7015 8117: contig of 1103 bp in length
8118 8217: gap of unknown length
8218 10026: contig of 1809 bp in length
10027 10126: gap of unknown length
10127 11596: contig of 1470 bp in length
11597 11696: gap of unknown length
11697 13146: contig of 1450 bp in length
13147 13246: gap of unknown length
13247 14748: contig of 1502 bp in length
14749 14848: gap of unknown length
14849 16421: contig of 1573 bp in length
16422 16521: gap of unknown length
16522 18572: contig of 2051 bp in length
18573 18672: gap of unknown length
18673 20389: contig of 1717 bp in length
20390 20489: gap of unknown length
20490 21514: contig of 1025 bp in length
21515 21614: gap of unknown length
21615 22642: contig of 1028 bp in length
22643 22742: gap of unknown length
22743 23756: contig of 1014 bp in length
23757 23856: gap of unknown length
23857 24874: contig of 1018 bp in length
24875 24974: gap of unknown length
24975 26410: contig of 1436 bp in length
26411 26510: gap of unknown length
26511 28869: contig of 2359 bp in length
28870 28969: gap of unknown length
28970 30219: contig of 1250 bp in length
30220 30319: gap of unknown length
30320 32382: contig of 2063 bp in length
32383 32482: gap of unknown length
32483 34188: contig of 1706 bp in length
34189 34288: gap of unknown length
34289 36188: contig of 1900 bp in length
36189 36289: gap of unknown length
36289 38198: contig of 1910 bp in length
38199 38298: gap of unknown length

38299 39999: contig of 1701 bp in length
40000 40099: gap of unknown length
40100 41725: contig of 1626 bp in length
41726 41825: gap of unknown length
41826 43012: contig of 1187 bp in length
43013 43112: gap of unknown length
43113 44380: contig of 1268 bp in length
44381 44480: gap of unknown length
44481 46394: contig of 1914 bp in length
46395 46494: gap of unknown length
46495 47880: contig of 1386 bp in length
47881 47980: gap of unknown length
47981 50212: contig of 2232 bp in length
50213 50312: gap of unknown length
50313 52820: contig of 2508 bp in length
52821 52920: gap of unknown length
52921 54278: contig of 1358 bp in length
54279 54378: gap of unknown length
54379 55889: contig of 1511 bp in length
55890 55989: gap of unknown length
55990 57668: contig of 1679 bp in length
57669 57769: gap of unknown length
60061 60160: gap of unknown length
60161 62005: contig of 1845 bp in length
62006 62105: gap of unknown length
62106 63410: contig of 1305 bp in length
63411 63510: gap of unknown length
63511 64895: contig of 1385 bp in length
64896 64996: gap of unknown length
64996 67080: contig of 2085 bp in length
67081 67180: gap of unknown length
67181 68364: contig of 1184 bp in length
68365 68464: gap of unknown length
68465 71000: contig of 2536 bp in length
71001 71100: gap of unknown length
71101 72970: contig of 1870 bp in length
72971 73070: gap of unknown length
73071 74831: contig of 1761 bp in length
74832 74931: gap of unknown length
74932 76508: contig of 1577 bp in length
76509 76608: gap of unknown length
76609 78985: contig of 2377 bp in length
78986 79085: gap of unknown length
79086 81746: contig of 2661 bp in length
81747 81846: gap of unknown length
81847 84951: contig of 3105 bp in length
84952 85051: gap of unknown length
85052 87753: contig of 2702 bp in length
87754 87853: gap of unknown length
87854 90939: contig of 3086 bp in length
90940 91039: gap of unknown length
91040 94370: contig of 3331 bp in length
94371 94470: gap of unknown length
94471 96108: contig of 1638 bp in length
96109 96208: gap of unknown length
96209 98946: contig of 2738 bp in length
98947 99046: gap of unknown length
99047 101632: contig of 2586 bp in length

Query Match
Best Local Similarity 100.0%; Score 15; DB 2; Length 108610;
Matches 15; Conservativity 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
Db 9039 AGCCATAACATGAAC 9053
|||||

RESULT 10
HS135L22/c 121212 bp DNA linear PRI 17-MAR-2000
LOCUS HS135L22
DEFINITION Human DNA sequence from clone RPL-135L22 on chromosome 6p22.2-22.3.

ACCESSION Contains STS and GSSs, complete sequence.
AL031767
VERSION AL031767.13 GI:6969156
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 121212)
REFERENCE Smalley,C.
AUTHORS Direct Submission
TITLE Submitted (13-MAR-2000) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 15A, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Feb 11, 2000 this sequence version replaced gi:6911919.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RPI-135L22 is from the library RPCI-1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: PCYPAC2
This sequence is the entire insert of clone RPI-135L22.
FEATURES
source
1. .121212
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p22.2-22.3"
/clone="RPI-135L22"
/clone_lib="RPCI-1"
578. .853
/note="AluX repeat: matches 21. .286 of consensus"
repeat_region
899. .1064
/note="L1MB2 repeat: matches 6001. .6166 of consensus"
repeat_region
2374. .2639
/note="AluX repeat: matches 1. .280 of consensus"
repeat_region
2976. .4675
/note="L1PA7 repeat: matches 4444. .6143 of consensus"
repeat_region
4748. .4900
/note="MIR repeat: matches 46. .221 of consensus"
repeat_region
4975. .5128
/note="MER63 repeat: matches 671. .1681 of consensus"
repeat_region
5388. .5871
/note="MER39 repeat: matches 54. .543 of consensus"
misc_feature
5530. .6001
/note="match: GSS: Em:B48646"
5947. .6269
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region
6994. .7284
/note="AluX repeat: matches 1. .291 of consensus"
repeat_region
8006. .8317
/note="AluSp repeat: matches 1. .313 of consensus"
repeat_region
8425. .8721
/note="AluX repeat: matches 1. .306 of consensus"

repeat_region 9185. .9646
/note="Tigger3b repeat: matches 757. .1229 of consensus"
repeat_region 9648. .9689
/note="21 copies 2 mer to 78% conserved"
repeat_region 9738. .10018
/note="AluX repeat: matches 1. .279 of consensus"
repeat_region 10019. .10443
/note="Tigger3b repeat: matches 351. .767 of consensus"
repeat_region 10444. .10739
/note="AluSc repeat: matches 1. .296 of consensus"
repeat_region 10740. .11066
/note="Tigger3b repeat: matches 1. .351 of consensus"
repeat_region 11105. .11249
/note="L2 repeat: matches 2605. .2750 of consensus"
misc_feature 11257. .11713
/note="match: GSS: Em:AQ187585"
11279. .11639
/note="LMC/D repeat: matches 5366. .5739 of consensus"
repeat_region 11730. .12003
/note="LTR23 repeat: matches 1. .278 of consensus"
repeat_region 12004. .12299
/note="AluX repeat: matches 1. .294 of consensus"
repeat_region 12300. .12326
/note="LTR23 repeat: matches 278. .303 of consensus"
repeat_region 12332. .12630
/note="AluSp repeat: matches 1. .302 of consensus"
repeat_region 12647. .12953
/note="AluX repeat: matches 1. .305 of consensus"
repeat_region 12957. .13088
/note="LTR23 repeat: matches 299. .429 of consensus"
repeat_region 13175. .13601
/note="MER39 repeat: matches 13. .477 of consensus"
repeat_region 14084. .14389
/note="AluSq repeat: matches 1. .305 of consensus"
repeat_region 14477. .14767
/note="AluSc repeat: matches 1. .290 of consensus"
repeat_region 14864. .14999
/note="MIR repeat: matches 109. .244 of consensus"
repeat_region 16325. .16543
/note="AluY repeat: matches 94. .311 of consensus"
repeat_region 17085. .17198
/note="MIR repeat: matches 70. .186 of consensus"
repeat_region 18083. .18283
/note="AluDb repeat: matches 96. .296 of consensus"
repeat_region 19807. .19850
/note="22 copies 2 mer to 77% conserved"
repeat_region 19980. .20284
/note="AluY repeat: matches 1. .304 of consensus"
repeat_region 20790. .21099
/note="MLT1A2 repeat: matches 3. .312 of consensus"
repeat_region 21109. .21302
/note="97 copies 2 mer to 57% conserved"
repeat_region 21445. .21719
/note="AluJo repeat: matches 13. .295 of consensus"
repeat_region 21720. .21791
/note="MLT1A2 repeat: matches 323. .394 of consensus"
repeat_region 22083. .22510
/note="MLT1C repeat: matches 6. .466 of consensus"
repeat_region 22671. .23511
/note="L1PB2 repeat: matches 5254. .6154 of consensus"
repeat_region 24714. .24745
/note="16 copies 2 mer aa 93% conserved"
repeat_region 25239. .25393
/note="MIR repeat: matches 28. .188 of consensus"
repeat_region 25468. .25690
/note="MER58A repeat: matches 1. .221 of consensus"
repeat_region 27578. .27688
/note="MLT1B repeat: matches 268. .390 of consensus"
repeat_region 27694. .28261
/note="L1PA14 repeat: matches 5590. .6149 of consensus"
repeat_region 28538. .30098
/note="L1PA7 repeat: matches 4586. .6141 of consensus"
repeat_region 30256. .30567

repeat_region	/note="AluSg repeat: matches 1. .310 of consensus" 30575. .30905
repeat_region	/note="MLT1C repeat: matches 3. .362 of consensus" 31496. .31796
repeat_region	/note="AluSg repeat: matches 1. .304 of consensus" 31812. .31903
repeat_region	/note="46 copies 2 mer ga 68% conserved" 33683. .33990
repeat_region	/note="AluIo repeat: matches 1. .303 of consensus" 34992. .35139
repeat_region	/note="MER63A repeat: matches 1. .177 of consensus" 35345. .35657
repeat_region	/note="AluSx repeat: matches 1. .312 of consensus" 36572. .36872
repeat_region	/note="AluSc repeat: matches 1. .299 of consensus" 37109. .37327
repeat_region	/note="L1MC1 repeat: matches 6113. .6325 of consensus" 37328. .37613
repeat_region	/note="AluY repeat: matches 1. .293 of consensus" 37860. .38192
misc_feature	/note="AluSx repeat: matches 1. .302 of consensus" 38805. .39089
repeat_region	/note="match: STS: Em:G02993" 39255. .39682
repeat_region	/note="LTR32 repeat: matches 30. .468 of consensus" 39701. .39902
repeat_region	/note="HERVL32 repeat: matches 3215. .3417 of consensus" 40020. .40309
repeat_region	/note="AluDb repeat: matches 1. .291 of consensus" 42183. .42494
repeat_region	/note="AluY repeat: matches 1. .311 of consensus" 42511. .42807
repeat_region	/note="AluSg repeat: matches 1. .297 of consensus" 42812. .45020
repeat_region	/note="HERVL32 repeat: matches 1274. .3213 of consensus" 45205. .45366
repeat_region	/note="MER5B repeat: matches 18. .178 of consensus" 45369. .45819
repeat_region	/note="LTR7 repeat: matches 1. .448 of consensus" 45820. .46198
repeat_region	/note="HERVH repeat: matches 7322. .7713 of consensus" 46197. .46413
repeat_region	/note="HERVH repeat: matches 5600. .5819 of consensus" 46415. .47075
repeat_region	/note="HERVH repeat: matches 4488. .5149 of consensus" 47078. .47582
repeat_region	/note="HERVH repeat: matches 3492. .3993 of consensus" 47573. .47722
repeat_region	/note="HERVH repeat: matches 3140. .3290 of consensus" 47702. .50691
repeat_region	/note="HERVH repeat: matches 1. .2995 of consensus" 50694. .51134

Query Match	100.0%;	Score 15;	DB 9;	Length 121212;
Best Local Similarity	100.0%;	Pred. No. 2e+02;		
Matches 15; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 AGCCATAACATGAC 15
 |||
 Db 66148 AGCCATAACATGAC 66134

RESULT	11				
AC131121/c					
LOCUS					
DEFINITION	AC131121	127400 bp	DNA	linear	HTG 16-AUG-2002
	Mus musculus	chromosome UNK clone	RP24-550I15,	WORKING	DRAFT
	SEQUENCE,	9 unordered	pieces.		
ACCESSION	AC131121				
VERSION	AC131121.1	GI:22267927			
KEYWORDS	HTG; HTGS_PHASE1;	HTGS_DRAFT.			
SOURCE	house mouse.				
ORGANISM	Mus musculus				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				

REFERENCE	1 (bases 1 to 127400)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	The sequence of Mus musculus clone
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 127400)
AUTHORS	McPherson,J.D. and Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (16-AUG-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watsn.wustl.edu
----- Project Information -----
Center project name: M_BB0550I15
----- Summary Statistics -----

```

Summary Statistics

Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 121880 bases at least Q40
Consensus quality: 123345 bases at least Q30
Consensus quality: 124257 bases at least Q20
Insert size: 188000; agarose-fp
Insert size: 126600; sum-of-contigs
Quality coverage: 4.12 in Q20 bases; agarose-fp
Quality coverage: 4.82 in Q20 bases; sum-of-contigs

quality coverage: 4.82 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* be preserved.

*	1	9874:	contig of 9874 bp in length
*	9875	9974:	gap of unknown length
*	9975	20688:	contig of 10714 bp in length
*	20689	20788:	gap of unknown length
*	20789	35406:	contig of 14618 bp in length
*	35407	35506:	gap of unknown length
*	35507	64851:	contig of 29345 bp in length
*	64852	64951:	gap of unknown length
*	64952	113805:	contig of 48854 bp in length
*	113806	113905:	gap of unknown length
*	113906	115892:	contig of 1987 bp in length
*	115893	115992:	gap of unknown length
*	115993	117914:	contig of 1922 bp in length
*	117915	118014:	gap of unknown length
*	118015	122126:	contig of 4112 bp in length
*	122127	122226:	gap of unknown length
*	122227	127400:	contig of 5174 bp in length.

FEATURES

Location/Qualifiers

Source

```
misc_feature      1. .9874  
                  /note="assembly_name:Contig10"  
misc_feature      9975. .20688  
                  /note="assembly_name:Contig11"  
misc_feature      20789. .35406  
                  /note="assembly_name:Contig12"  
misc_feature      35507. .64851  
                  /note="assembly_name:Contig13"  
misc_feature      64952. .113805
```


misc_feature /note="assembly_name:Contig14"
113906.115892
misc_feature /note="assembly_name:Contig6"
115993.117914
/note="assembly_name:Contig7
clone_end:Sp6
vector_side:right"
misc_feature 118015.122126
/note="assembly_name:Contig8"
misc_feature 122227.127400
/note="assembly_name:Contig9"
BASE COUNT 36146 a 27561 c 26883 g 36009 t 801 others
ORIGIN
Query Match 100.0%; Score 15; DB 2; Length 127400;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCATAACATGAAC 15
|||||
Db 63666 AGCCATAACATGAAC 63652
RESULT 12
AC006434 129547 bp DNA linear PLN 25-JUL-2000
LOCUS
DEFINITION Genomic sequence for Arabidopsis thaliana BAC F10A5, complete
sequence.
AC006434
AC006434.5 GI:4662609
HTG.
Arabidopsis thaliana.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 129547)
Ecker,J.R.
Direct Submission
Submitted (24-JAN-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
2 (bases 1 to 129547)
Ecker,J.R.
Direct Submission
Submitted (23-APR-1999) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th Street and
Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
3 (bases 1 to 129547)
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (28-JUN-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
4 (bases 1 to 129547)
Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
Direct Submission
Submitted (22-JUL-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
5 (bases 1 to 129547)

AUTHORS Cheuk,R., Shinn,P., Brooks,S., Buehler,E., Chao,Q.,
Johnson-Hopson,C., Khan,S., Kim,C., Altafi,H., Bei,B., Chin,C.,
Chiou,J., Choi,E., Conn,L., Conway,A., Gonzalez,A., Hansen,N.,
Howing,B., Koo,T., Lam,B., Lee,J., Lenz,C., Li,J., Liu,A., Liu,J.,
Liu,S., Mukharsky,N., Nguyen,M., Palm,C., Pham,P., Sakano,H.,
Schwartz,J., Southwick,A., Thaveri,A., Toriumi,M., Vaysberg,M.,
Yu,G., Davis,R., Federspiel,N., Theologis,A. and Ecker,J.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2000) Arabidopsis thaliana Genome Center,
Department of Biology, University of Pennsylvania, 38th and
Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT On Apr 23, 1999 this sequence version replaced gi:4416544.
FEATURES
source Location/Qualifiers
1.129547
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(join(<1..290,372..426))
/note="similar to receptor serine/threonine kinase-like
protein emb|CA16797.1"
/codon_start=1
/evidence=not_experimental
/product="F10A5.1"
/protein_id="AAF87135.1"
/db_xref="GI:9369386"
/translation="MALPLPLIFLIFSHLIVSGVRSSTFIWVNCCEYTWPGLLSNAG
VPPLPTGIFVLQKGEERTISAPTSWGRFWGRGTQCTDTRKFTCLTGDCSGTLFES
GSGATPATLAEF"
join(1927..2064,2190..2255,2470..2643,2856..2958,
3056..3416,3520..3885,3968..4046,4126..4335)
/note="similar to L-ascorbate oxidase homolog precursor
sp|Q00624|ASO_BRANA"
/codon_start=1
/evidence=not_experimental
/product="F10A5.2"
/protein_id="AAF87105.1"
/db_xref="GI:9369356"
/translation="MRHVEVEVLVLSVLELSYAFAPISISYQWVVSYSQRFILGN
KQVIVINDMFPPIINATANDIIVKQDIGSYFYFPTLLLOKAAGYGAIRIYPPELV
PVPRPKDEEYDILIGDWFYLDHTVMRASLDAGSHLPNPGILFNGRPEETFAPER
GKYRRLRISNVGLKTCLEFRIDQDMLLVETEGTYQKRVYSSLDIHGQSYSLVTA
KTPDVGIRSYIIFATARETDSLGIALIRYGPSPLDPVGQPLAPALQDFGSSVEQ
ALSIRMDLVNGAARSNPQGSYHYGRINVTRTIILHNDVMSGKLRYTINGVSFVPE
TPKLVDHFQLNDTIIPGMFVYPSNKTPTLIGTSVDIHYKDFIHVFONPLFGLESY
HIDGYNEFVVGYFGGAWSESCKAGYNLVDVSRSTVOYYPYSWTAILAMDNOGMNV
RSQAEQWYVLGQELMYRVKGEEDPSTIPVRDENPIFGNVIRCGKVR"
join(5713..6109,6903..7172,7261..7937)
/note="putative tubulin beta-4 chain sp|P50262|TBB4_PORPU;
similar to ESTs dbj|AV554594.1, dbj|AV536568.1,
dbj|AV543492.1, dbj|AV542234.1, and gb|W43629.1"
/codon_start=1
/evidence=not_experimental
/product="F10A5.3"
/protein_id="AAF87106.1"
/db_xref="GI:9369357"
/translation="MREILHVQGGQCGNQIGSKFWEYICDEHGVDPTRYNQDSADLQ
LERINVYNEASGGRYVPRAVLMDLEPGTMDSIRSGPYGQIFRPDNFVFGQSGAGNMW
AKGHYTEGAELIDAVLDVVRKEAENCDCLOGFYVCHSLGGGTSGMGTLISKIREY
PDRMLTFSVPSPKYSVTVEPYNATLSVHQLVENADECMVLNDNEALYDICFRTLKL
STPSFGDLNHLISATMSGVTCISLRFPGQLNSDLKLAVNLIPRRLHFFWVGAPLTS
RGSQYISLTVPETLQMDKAMNMCAADPRHGRYLTASAMEFGKMSKEVDILNV
QNKSSYFVEMIPNNVKSVCIDIPPTGIKMASTFVGNSTSIQEMFRVSEQFTAMFRR
KAFLHWYTGEGMDEMEFTEAESNMNDLVSEYQYQODATADEEDYEDEEEQYVES"
complement(join(8344..8471,8704..8782,8967..9293))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F10A5.4"
/protein_id="AAF87134.1"
/db_xref="GI:9369385"
/translation="MSKKKKTOSLSMEDHRDGKGRKDDFFDSIDYEISSILLQLSHP
VIFSSDSPVLHKWGRKRRSSSSHLHPPEVYTKSLPCMAVDGDMGSSSSSCLTGEA
KKTNSQIEKSEFFSPRGVHCYLIIFCFPPCTIKSSEEGFORVPNKVQVAQARQLGL

IRSKKLFNRFISYQFK"
join(10894. .11116,11285. .11484,11708. .11848,11931. .12002,
12074. .12198,12327. .12375)
/note="unknown protein; similar to EST dbj|AV559983.1"
/codon_start=1
/evidence=not_experimental
/product="F10A5.5"
/protein_id="AAF87107.1"
/db_xref="GI:9369358"
/translation="MKATRPPIHAVTTWRQRPKYKGLGVSAMTALVLRIVHD
HDNLFVAAEAHVSLGISVLIKLTKERTCAGLSKSQELTALFLAVRLYCSFMEEDI
HTLLDSATLVTTLFVIYIMIRFKLKASYMDKDNFAIYVVPVLSVLHPSTHHI
INKISWAFCVYLEAVSLPOLRVMQNTKPTAHYVFAIGIARFLSCAHWLVQVLDTRG
RLTLALGYGFPIWVLSLSEIVQTEFILADFCYIYVKSLSMGQVLRLRPSGVV"
join(17031. .17238,17321. .17502)
/note="similar to GAST-like gene product gb|AAB97006.1;
similar to ESTs dbj|AV536142.1, and dbj|AV566951.1"
/codon_start=1
/evidence=not_experimental
/product="F10A5.6"
/protein_id="AAF87108.1"
/db_xref="GI:9369359"
/translation="MAISKALIASLLISLVQLVQADVRLFHKKLIILNIILMFSK
TYFSHTTIFLQENSQKKNGYAKKIDCGSACVARCRLSRPRLCHRACGTCYRCNCV
PGTYGNYDKQCQYASLTFHGGRRKCP"
complement(18355. .18426)
product="tRNA-Thr"
/note="codon recognized: ACG; putative tRNA-Thr"
/evidence=not_experimental
join(20514. .20675,20808. .20865,21251. .21363,21464. .21753,
21840. .21900,21989. .22025,22094. .22157,22244. .22643,
22721. .22852,22938. .23324)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F10A5.7"
/protein_id="AAF87109.1"
/db_xref="GI:9369360"
/translation="MDLGSGETSKSKGMLKLGFADEKLKQENSYDTLPSPDLENV
SKKSKLPKKNLKDNTGVDHASVPRLKRSAMKRNLESVSKSSVRLNRFKTGES
SVKESQEMEAKAIVTESMISDEKEVAETLYGLAGMFTETDSLKKTSEKETSKV
VDSLIVEDDYTKTESLTPVSVLSSAKTKQIDAMPLEQSDKQFNTGMVDFIDRLKQ
SSSVKVTDAHARVNETKVTTSNGALWPLGSLSTVHSGAQLSKPSTKLPWMGQAV
SPSNTASLNLNCEPLRVQPRKLRKACASHIYISRLIKVLQTSKSLATLNQNEQRYSEMS
ERRLPDSVITINDFKAIVSPKRPQNPHLDMHRAHNPVPQEDMTKLALELYGPHTS
OKQSFDFLSSAGAAQSHFPLNSFPQYQISAAYNSQLSPATSSHQQLSPYLASRFQ
TAYNANQQQQOQLOKRLWAAQYRPTNGNTMQSNQYSKPNLSLNTSIQQPLOWSSPR
YNHNVSQQQHRRLMTAAAAAMSMSHQHNNPSRAVMNRQEHHPFLIYEDTRTPQLLCNEQ
S"
join(23997. .24116,24212. .24667)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F10A5.8"
/protein_id="AAF87110.1"
/db_xref="GI:9369361"
/translation="METTKRAEIDTTAPRTVKEAVALFGERVLASQVYSNHLKMHDE
KWEDPSGIKIELQETRYDLKRAKESIQMRNSLSCLKELERTKQELQKLRVDPGVNE
TKLDETVEKTFEVLVPRVDDPEIRSPRLRSMSEKRYVKFANPTGNNGSVFLERHPSM
KKKEKTKDKKKSLIPLFIGIGIFSKKKVLQ"
complement(25054. .25338)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F10A5.9"
/protein_id="AAF87113.1"
/db_xref="GI:9369384"
/translation="MKLFHKLGIKLMRLIDFTLHSSSRQKKNSRSGDGERFEPPKI
SCSSSYSSHVHYSEAIADCIIEFNKSSAMSCEDHEARLVNNIDRCYV"
complement(join(28187. .28890,29737. .29834,29923. .30509))
/note="unknown protein; similar to ESTs gb|BE038339.1,
gb|R30294.1, dbj|AV545596.1, and gb|AI992442.1"
/codon_start=1

/evidence=not_experimental
/product="F10A5.10"
/protein_id="AAF87132.1"
/db_xref="GI:9369383"
/translation="MALITFLPENAEAPQKHKPTSSKRRKRDNPDTQTQTKHKPKP
KKAVPKPQSSWDQIKNLITCKQIEGSRVHDPKSNQSGPSMTNLSPSKLGSSCSSI

Query Match	100.0%;	Score 15;	DB 8;	Length 129547;
Best Local Similarity	100.0%;	Pred. No. 2e+02;		
Matches	15;	Conservative 0;	Mismatches 0;	Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
|||||
Db 113975 AGCCATAACATGAAC 113989

RESULT 13	AC016036/c	145520 bp	DNA	linear	HTG 13-JUL-2000
LOCUS	Homo sapiens clone RP11-23K12, LOW-PASS SEQUENCE SAMPLING.				
DEFINITION	AC016036				
ACCESSION	AC016036.3	GI:9128702			
VERSION	HTG; HTGS_PHASE0.				
KEYWORDS	Homo sapiens.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 145520)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C. and Lander,E.				
JOURNAL	Homo sapiens, clone RP11-23K12				
REFERENCE	2 (bases 1 to 145520)				
AUTHORS	Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.				
TITLE	Direct Submission				
JOURNAL	Submitted (19-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA				
COMMENT	On Jul 13, 2000 this sequence version replaced gi:7108033. All repeats were identified using RepeatMasker: Smit,A.F.A. & Green,P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html ----- Genome Center Center: Whitehead Institute/ MIT Center for Genome Research Center code: WIBR Web site: http://www-seq.wi.mit.edu Contact: sequence_submissions@genome.wi.mit.edu ----- Project Information Center project name: L4386 Center clone name: 23_K_12 ----- * NOTE: This record contains 159 individual * sequencing reads that have not been assembled into * contigs. Runs of N are used to separate the reads * and the order in which they appear is completely * arbitrary. Low-pass sequence sampling is useful for * identifying clones that may be gene-rich and allows * overlap relationships among clones to be deduced. * However, it should not be assumed that this clone * will be sequenced to completion. In the event that * the record is updated, the accession number will * be preserved.				

1	715: contig of 715 bp in length	30894	30993: gap of 100 bp
716	815: gap of 100 bp	30994	31729: contig of 736 bp in length
816	1545: contig of 730 bp in length	31730	31829: gap of 100 bp
1546	1645: gap of 100 bp	31830	32540: contig of 711 bp in length
1646	2357: contig of 712 bp in length	32541	32640: gap of 100 bp
2358	2457: gap of 100 bp	32641	33372: contig of 732 bp in length
2458	3201: contig of 744 bp in length	33373	33472: gap of 100 bp
3202	3301: gap of 100 bp	33473	34203: contig of 731 bp in length
3302	4041: contig of 740 bp in length	34204	34303: gap of 100 bp
4042	4141: gap of 100 bp	34304	35038: contig of 735 bp in length
4142	4900: contig of 759 bp in length	35039	35138: gap of 100 bp
4901	5000: gap of 100 bp	35139	35875: contig of 737 bp in length
5001	5733: contig of 733 bp in length	35876	35975: gap of 100 bp
5734	5833: gap of 100 bp	35976	36711: contig of 736 bp in length
5834	6556: contig of 723 bp in length	36712	36811: gap of 100 bp
6557	6656: gap of 100 bp	36812	37547: contig of 736 bp in length
7378	7377: contig of 721 bp in length	37548	37647: gap of 100 bp
7378	7477: gap of 100 bp	37648	38408: contig of 761 bp in length
7478	8204: contig of 727 bp in length	38409	38508: gap of 100 bp
8205	8304: gap of 100 bp	38509	39244: contig of 736 bp in length
8305	9029: contig of 725 bp in length	39245	39344: gap of 100 bp
9030	9129: gap of 100 bp	39345	40060: contig of 716 bp in length
9130	9867: contig of 738 bp in length	40061	40160: gap of 100 bp
9868	9967: gap of 100 bp	40161	40893: contig of 733 bp in length
9968	10712: contig of 745 bp in length	40894	40993: gap of 100 bp
10713	10812: gap of 100 bp	40994	41721: contig of 728 bp in length
10813	11550: contig of 738 bp in length	41722	41821: gap of 100 bp
11551	11650: gap of 100 bp	41822	42570: contig of 749 bp in length
11651	12407: contig of 757 bp in length	42571	42670: gap of 100 bp
12408	12507: gap of 100 bp	42671	43409: contig of 739 bp in length
12508	13264: contig of 757 bp in length	43410	43509: gap of 100 bp
13265	13364: gap of 100 bp	43510	44264: contig of 755 bp in length
13365	14107: contig of 743 bp in length	44265	44364: gap of 100 bp
14108	14207: gap of 100 bp	44365	45101: contig of 737 bp in length
14208	14932: contig of 725 bp in length	45102	45201: gap of 100 bp
14933	15032: gap of 100 bp	45202	45947: contig of 746 bp in length
15033	15749: contig of 717 bp in length	45948	46047: gap of 100 bp
15750	15849: gap of 100 bp	46048	46791: contig of 744 bp in length
15850	16589: contig of 740 bp in length	46792	46891: gap of 100 bp
16590	16689: gap of 100 bp	46892	47619: contig of 728 bp in length
16690	17435: contig of 746 bp in length	47620	47719: gap of 100 bp
17436	17535: gap of 100 bp	47720	48444: contig of 725 bp in length
17536	18279: contig of 744 bp in length	48445	48544: gap of 100 bp
18280	18379: gap of 100 bp	48545	49273: contig of 729 bp in length
18380	19125: contig of 746 bp in length	49274	49373: gap of 100 bp
19126	19225: gap of 100 bp	49374	50076: contig of 703 bp in length
19226	19984: contig of 759 bp in length	50077	50176: gap of 100 bp
19985	20084: gap of 100 bp	50177	50923: contig of 747 bp in length
20085	20826: contig of 742 bp in length	50924	51023: gap of 100 bp
20827	20926: gap of 100 bp	51024	51752: contig of 729 bp in length
20927	21666: contig of 740 bp in length	51753	51852: gap of 100 bp
21667	21766: gap of 100 bp	51853	52601: contig of 749 bp in length
21767	22511: contig of 745 bp in length	52602	52701: gap of 100 bp
22512	22611: gap of 100 bp	52702	53449: contig of 748 bp in length
22612	23333: contig of 722 bp in length	53449	53549: gap of 100 bp
23334	23433: gap of 100 bp	53550	54302: contig of 753 bp in length
23434	24172: contig of 739 bp in length	54303	54402: gap of 100 bp
24173	24272: gap of 100 bp	54403	55147: contig of 745 bp in length
24273	25022: contig of 750 bp in length	55148	55247: gap of 100 bp
25023	25122: gap of 100 bp	55248	56008: contig of 761 bp in length
25123	25847: contig of 725 bp in length	56009	56108: gap of 100 bp
25848	25947: gap of 100 bp	56109	56847: contig of 739 bp in length
25948	26700: contig of 753 bp in length	56848	56947: gap of 100 bp
26701	26800: gap of 100 bp	56948	57684: contig of 737 bp in length
26801	27522: contig of 722 bp in length	57685	57784: gap of 100 bp
27523	27622: gap of 100 bp	57785	58536: contig of 752


```

repeat_region      14168..14517
                    /rpt_family="MALR"
repeat_region      14531..14747
                    /rpt_family="L1"
repeat_region      14903..16651
                    /rpt_family="L1"
repeat_region      16655..16914
                    /rpt_family="Alu"
repeat_region      16952..17403
                    /rpt_family="L1"
repeat_region      17415..17609
                    /rpt_family="L1"
repeat_region      18536..18614
                    /rpt_family="MIR"
repeat_region      20111..20186
                    /rpt_family="MIR"
repeat_region      20237..20516
                    /rpt_family="MIR"
repeat_region      20867..21289
                    /rpt_family="L2"
repeat_region      23662..23850
                    /rpt_family="MERL_type"
repeat_region      23864..23960
                    /rpt_family="Alu"
repeat_region      23984..24136
                    /rpt_family="MALR"
repeat_region      25154..25483
                    /rpt_family="L1"
repeat_region      25563..25610
                    /rpt_family="MERL_type"
repeat_region      26366..26464
                    /rpt_family="MERL_type"
repeat_region      26631..26829
                    /rpt_family="MIR"
repeat_region      28319..28616
                    /rpt_family="Alu"
repeat_region      29095..29373
                    /rpt_family="Alu"
repeat_region      29391..29660
                    /rpt_family="Alu"
repeat_region      30983..31066
                    /rpt_family="MIR"
repeat_region      31668..32325
                    /rpt_family="L2"
repeat_region      32326..32638
                    /rpt_family="L1"
repeat_region      32639..32812
                    /rpt_family="Alu"
repeat_region      32830..33552
                    /rpt_family="L1"
repeat_region      33597..33964
                    /rpt_family="L1"
repeat_region      34022..34096
                    /rpt_family="L1"
repeat_region      34097..34392
                    /rpt_family="L1"
repeat_region      34393..35687
                    /rpt_family="Alu"
repeat_region      35818..36495
                    /rpt_family="L1"
repeat_region      36546..36601
                    /rpt_family="L1"
repeat_region      36654..38091
                    /rpt_family="L1"
repeat_region      38084..39798
                    /rpt_family="L1"

```

Query Match 100.0%; Score 15; DB 9; Length 154084;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
 |||||
 Db 91395 AGCCATAACATGAAC 91409

RESULT 15
 AC130009/c
 LOCUS
 DEFINITION
 AC130009
 158861 bp DNA linear HTG 08-AUG-2002
 Rattus norvegicus clone CH230-324C5, *** SEQUENCING IN PROGRESS
 *, 55 unordered pieces.
 AC130009
 AC130009.1 GI:22138172
 HTG: HTGS_PHASE1.
 Norway rat.
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE
 1 (bases 1 to 158861)
 Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
 Barbara,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
 Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
 Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
 Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
 Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
 Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
 Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
 Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
 Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
 Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
 Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
 Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louisedge,H.,
 Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
 Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
 Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
 Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
 Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,S., Ren,Y.,
 Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
 Scherer,S., Scott,G., Shen,H., Shooshitari,N., Sisson,I.,
 Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
 Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
 Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
 Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 Weinstein,G., and Gibbs,R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 Project Information
 Center project name: KATP
 Center clone name: CH230-324C5
 Summary Statistics

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 124544 bases at least Q40
Consensus quality: 128196 bases at least Q30
Consensus quality: 131011 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 2609: contig of 1307 bp in length
* 2610 2709: gap of unknown length
* 2710 3850: contig of 1141 bp in length
* 3851 3950: gap of unknown length
* 3951 4978: contig of 1028 bp in length
* 4979 5078: gap of unknown length
* 5079 6269: contig of 1191 bp in length
* 6270 6369: gap of unknown length
* 6370 7473: contig of 1104 bp in length
* 7474 7573: gap of unknown length
* 7574 8907: contig of 1334 bp in length
* 8908 9007: gap of unknown length
* 9008 10051: contig of 1044 bp in length
* 10052 10151: gap of unknown length
* 10152 11191: contig of 1040 bp in length
* 11192 11291: gap of unknown length
* 11292 12427: contig of 1136 bp in length
* 12428 12527: gap of unknown length
* 12528 13722: contig of 1195 bp in length
* 13723 13822: gap of unknown length
* 13823 15030: contig of 1208 bp in length
* 15031 15130: gap of unknown length
* 15131 17569: contig of 2439 bp in length
* 17570 17669: gap of unknown length
* 17670 19479: contig of 1810 bp in length
* 19480 19579: gap of unknown length
* 19580 20965: contig of 1386 bp in length
* 20966 21065: gap of unknown length
* 21066 22711: contig of 1646 bp in length
* 22712 22811: gap of unknown length
* 22812 24497: contig of 1686 bp in length
* 24498 24597: gap of unknown length
* 24598 26106: contig of 1509 bp in length
* 26107 26206: gap of unknown length
* 26207 27767: contig of 1561 bp in length
* 27768 27867: gap of unknown length
* 27868 28973: contig of 1106 bp in length
* 28974 29073: gap of unknown length
* 29074 30550: contig of 1477 bp in length
* 30551 30650: gap of unknown length
* 30651 32496: contig of 1846 bp in length
* 32497 32596: gap of unknown length
* 32597 33896: contig of 1300 bp in length
* 33897 33996: gap of unknown length
* 33997 35948: contig of 1952 bp in length
* 35949 36048: gap of unknown length
* 36049 37884: contig of 1836 bp in length
* 37885 37984: gap of unknown length
* 37985 39844: contig of 1860 bp in length
* 39845 39944: gap of unknown length
* 39945 41441: contig of 1497 bp in length
* 41442 41541: gap of unknown length
* 41542 42674: contig of 1133 bp in length
* 42675 42774: gap of unknown length

* 42775 44391: contig of 1617 bp in length
* 44392 44491: gap of unknown length
* 44492 46237: contig of 1746 bp in length
* 46238 46337: gap of unknown length
* 46338 47773: contig of 1436 bp in length
* 47774 47873: gap of unknown length
* 47874 51167: contig of 3294 bp in length
* 51168 51267: gap of unknown length
* 51268 53726: contig of 2459 bp in length
* 53727 53826: gap of unknown length
* 53827 56626: contig of 2800 bp in length
* 56627 56726: gap of unknown length
* 56727 60004: contig of 3278 bp in length
* 60005 60104: gap of unknown length
* 60105 63599: contig of 3495 bp in length
* 63600 63699: gap of unknown length
* 63700 66017: contig of 2318 bp in length
* 66018 66117: gap of unknown length
* 66118 68050: contig of 1933 bp in length
* 68051 68150: gap of unknown length
* 68151 72358: contig of 4208 bp in length
* 72359 72458: gap of unknown length
* 72459 74954: contig of 2496 bp in length
* 74955 75054: gap of unknown length
* 75055 78265: contig of 3211 bp in length
* 78266 78365: gap of unknown length
* 78366 81407: contig of 3042 bp in length
* 81408 81507: gap of unknown length
* 81508 84643: contig of 3136 bp in length
* 84644 84743: gap of unknown length
* 84744 88452: contig of 3709 bp in length
* 88453 88552: gap of unknown length
* 88553 93253: contig of 4701 bp in length
* 93254 93353: gap of unknown length
* 93354 98211: contig of 4858 bp in length
* 98212 98312: gap of unknown length
* 98312 102965: contig of 4654 bp in length
* 102966 103065: gap of unknown length
* 103066 107149: contig of 3984 bp in length
* 107150 113981: contig of 6832 bp in length
* 113982 114081: gap of unknown length
* 114082 119381: contig of 5300 bp in length
* 119382 119481: gap of unknown length
* 119482 125815: contig of 6334 bp in length
* 125816 125915: gap of unknown length
* 125916 132747: contig of 6832 bp in length
* 132748 132847: gap of unknown length
* 132848 141339: contig of 8492 bp in length
* 141340 141439: gap of unknown length
* 141440 148269: contig of 6830 bp in length
* 148270 148369: gap of unknown length
* 148370 158861: contig of 10492 bp in length.

FEATURES
source 1. 158861
location/Qualifiers
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/clone="CH230-324C5"

Query Match 100.0%; Score 15; DB 2; length 158861;
Best local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
 |||||
Db 115471 AGCCATAACATGAAC 115457

Search completed: June 26, 2003, 03:47:37
Job time : 430.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(Without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	20	AAAX32405	Ab5 variable heavy
2	14	93.3	2401	23	ABL22930	Drosophila melanog
3	14	93.3	5235	23	ABL08064	Drosophila melanog
4	14	93.3	10425	23	ABL13480	Drosophila melanog
5	14	93.3	10426	23	ABL19870	Drosophila melanog
6	13.4	89.3	306	21	AAC56643	Eucalyptus grandis
7	13.4	89.3	348	21	AAC55981	Eucalyptus grandis
8	13.4	89.3	428	21	AAC54150	Arabidopsis thaliana
9	13.4	89.3	521	21	AAA43836	Chicken secreted e

C	10	13.4	89.3	556	22	ABA62712	Human foetal liver
C	11	13.4	89.3	556	22	ABA30013	Probe #8479 for ge
C	12	13.4	89.3	556	22	AAK11091	Human brain expres
C	13	13.4	89.3	556	22	AAK36915	Human bone marrow
C	14	13.4	89.3	556	22	AAI17764	Probe #7697 for ge
C	15	13.4	89.3	556	22	AAI42716	Probe #11402 used
C	16	13.4	89.3	556	24	ABS10923	Human genome-deriv
C	17	13.4	89.3	597	21	AAC34482	Arabidopsis thaliana
C	18	13.4	89.3	731	11	AAO05977	Gas vesicle proteol
C	19	13.4	89.3	1001	24	ABK51547	DNA upstream of Ha
C	20	13.4	89.3	1047	20	AAZ31559	Bacillus species p
C	21	13.4	89.3	1047	22	AAC86597	DNA encoding a Bac
C	22	13.4	89.3	1183	23	AAS59699	Proionibacterium
C	23	13.4	89.3	1416	22	AAK70918	Human immune/haema
C	24	13.4	89.3	1517	21	AAZ65319	Human secreted pro
C	25	13.4	89.3	1518	21	AAZ65348	Human secreted pro
C	26	13.4	89.3	1612	7	AAN60392	Sequence encoding
C	27	13.4	89.3	1685	24	ABN83054	Human development
C	28	13.4	89.3	1687	18	AAT84244	Human ICAM-4 gene
C	29	13.4	89.3	1687	19	AAV11684	Human ICAM-4 upstr
C	30	13.4	89.3	1719	21	AAI72859	Chicken GW75 cDNA.
C	31	13.4	89.3	2066	22	AAE27219	Human neovasculari
C	32	13.4	89.3	2181	21	AAZ99576	cDNA encoding a hu
C	33	13.4	89.3	2340	20	AAAX80866	DNA encoding human
C	34	13.4	89.3	2354	21	AAZ50348	Human secreted pro
C	35	13.4	89.3	2393	21	AAZ46135	cDNA encoding an a
C	36	13.4	89.3	2430	21	AAI72855	CDNA clone AR2, AT
C	37	13.4	89.3	2457	22	AAS09680	Oat cDNA encoding
C	38	13.4	89.3	3355	20	AAAX36342	Human TIE ligand N
C	39	13.4	89.3	3355	21	AAA77529	Human PRO188 cDNA
C	40	13.4	89.3	3355	22	AAS21382	Human cDNA sequenc
C	41	13.4	89.3	3355	22	AAC97379	Human angiogenesis
C	42	13.4	89.3	3355	24	AAD31564	Human TIE ligand N
C	43	13.4	89.3	6306	22	ABA15649	Human nervous syst
C	44	13.4	89.3	6695	22	ABA15648	Human nervous syst
C	45	13.4	89.3	6697	22	ABA15650	Human nervous syst

ALIGNMENTS

RESULT 1	AAAX32405	standard; DNA; 15 BP.
ID	AAAX32405	
XX		
AC	AAAX32405;	
XX		
DT	17-JUN-1999	(first entry)
XX		
DE	Ab5 variable heavy (VH) chain CDR1 encoding DNA.	
XX		
KW	Agonist antibody; thrombopoietin receptor; TPO-R; thrombopoietin; DIC;	
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;	
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;	
KW	myelodysplasia; myelotoxic chemotherapy; leukaemia; tumour; MusK; CDR;	
KW	neuromuscular; muscular dystrophy; complementarity determining region;	
KW	variable heavy chain; variable light chain; VH; VL; SS.	
XX		
OS	Homo sapiens.	
XX		
PN	WO9910494-A2.	
XX		
PD	04-MAR-1999.	
XX		
PF	21-AUG-1998;	98WO-US17364.
XX		
PR	25-AUG-1997;	97US-0918148.
XX		
PA	(GETH) GENENTECH INC.	
XX		
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;	
XX		
DR	WPI; 1999-204666/17.	

DR P-PSDB; AAY06704.
XX
PT New thrombopoietin receptor agonist antibodies - useful for
PT treating immunological or hematological disorders
XX
PS Claim 10; Page 80; 86pp; English.
XX
CC The invention relates to an agonist antibody (Ab) which binds to a
CC thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
CC be used in the same way and for the same indications as thrombopoietin
CC (TPO). They can stimulate proliferation, differentiation or growth of
CC megakaryocytes. They may also be able to stimulate megakaryocytes to
CC increase platelet production. They can be used for treating
CC immunological or hematopoietic disorders, especially thrombocytopenia.
CC Thrombocytopenia -associated bone marrow hypoplasia (e.g. aplastic anemia
CC following chemotherapy or bone marrow transplant) may be effectively
CC treated with the antibody compounds as well as disorders such as
CC disseminated intravascular coagulation (DIC), immune thrombocytopenia
CC (HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
CC congenital thrombocytopenia, thrombotic thrombocytopenia and
CC myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
CC treatment of solid tumours or leukaemia, myeloablative chemotherapy for
CC autologous or allogeneic bone marrow transplant, myelodysplasia,
CC idiopathic aplastic anemia, congenital thrombocytopenia, and immune
CC thrombocytopenia. The antibodies which bind to the MusK receptor can be
CC used for improving neuromuscular function in a patient, e.g. in muscular
CC dystrophy. The products can also be used for detection and diagnosis. The
CC antibodies have a longer half-life than the natural ligand for the TPO-R.
CC Sequences AAX32387-X32413 represent DNA fragments encoding the CDR1,
CC CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
CC (VL) chains of antibodies Ab1 to Ab6.
XX
SQ Sequence 15 BP; 7 A; 4 C; 2 G; 2 T; 0 other;

Query Match 100.0%; Score 15; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 96;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
Db 1 AGCCATAACATGAAC 15

RESULT 2
ABL22930
ID ABL22930 standard; DNA; 2401 BP.
XX
AC ABL22930;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 20263.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 20263; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2401 BP; 687 A; 464 C; 603 G; 647 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 2401;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCATAACATGAAC 15
Db 672 GCCATAACATGAAC 685

RESULT 3
ABL08064/C
ID ABL08064 standard; cDNA; 5235 BP.
XX
AC ABL08064;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 18674.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR P-PSDB; ABB63961.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 18674; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5235 BP; 1265 A; 1209 C; 1279 G; 1482 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 5235;
Best Local Similarity 100.0%; Pred. No. 4.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAA 14
|||||
DB 960 AGCCATAACATGAA 947

RESULT 4
ABL13480
ID ABL13480 standard; cDNA; 10425 BP.
XX
AC ABL13480;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34922.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR P-PSDB; ABB69377.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 34922; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 10425 BP; 2780 A; 2460 C; 2408 G; 2777 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 10425;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GCCATAACATGAAC 15
|||||

Db 4051 GCCATAACATGAAC 4064

RESULT 5
ABL19870
ID ABL19870 standard; DNA; 10426 BP.
XX
AC ABL19870;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 11083.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PD New isolated nucleic acid detection reagent for detecting 1000 or more
PD genes from Drosophila and for elucidating cell signalling and cell-cell
PD interactions -
XX
PS Claim 1; SEQ ID NO 11083; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 10426 BP; 2780 A; 2460 C; 2409 G; 2777 T; 0 other;

Query Match 93.3%; Score 14; DB 23; Length 10426;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GCCATAACATGAAC 15
|||||
Db 4051 GCCATAACATGAAC 4064

RESULT 6
AAC56643/C
ID AAC56643 standard; DNA; 306 BP.
XX
AC AAC56643;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #514.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

KW poplar; sweetgum; teak; mahogany; bzIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;
 KW homeobox zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200053724-A2.
 XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX
 DR WPI; 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 PS Claim 1; Page 482; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bzIP, bzIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeobox/MADS, homeobox zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 306 BP; 97 A; 49 C; 80 G; 80 T; 0 other;
 Query Match 89.3%; Score 13.4; DB 21; Length 306;
 Best Local Similarity 93.3%; Pred. No. 8.3e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGCCATACATGAC 15
 ID AACC55981 standard; DNA; 348 BP.
 XX
 AC AACC55981;
 XX
 DT 25-JAN-2001 (first entry)
 XX
 DE Eucalyptus grandis transcription factor DNA sequence #112.
 XX
 KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
 KW poplar; sweetgum; teak; mahogany; bzIP; G-box binding factor;
 KW basic helix-loop-helix zipper; homeotic; homeobox; MADS;
 KW homeobox zipper; LIM domain; AP2; ERBS; zinc finger domain;
 KW type 2 Cys2His2; CCAAT box element; MYB; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO200053724-A2.

XX
 PD 14-SEP-2000.
 XX
 PF 09-MAR-2000; 2000WO-US06112.
 XX
 PR 11-MAR-1999; 99US-0266513.
 PR 18-AUG-1999; 99US-0149485.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PI Wood M, McGrath A, Shenk MA, Glenn M;
 XX
 DR WPI; 2000-579369/54.
 XX
 PT New isolated polynucleotide encoding a plant transcription factor for
 PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
 PT having modified gene expression or modified activity of a polypeptide
 PT
 PS Claim 1; Page 73; 747pp; English.
 XX
 CC The present invention relates to novel plant transcription factors from
 CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
 CC sequence for one such transcription factor. The transcription factor may
 CC be used to produce a plant having modified gene expression such as a
 CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
 CC mahogany species or to modify the activity of a polypeptide in a plant.
 CC The transcription factors of the present invention are members from the
 CC following families of regulatory proteins: bzIP, bzIP family of G-box
 CC binding factors, basic helix-loop-helix zipper,
 CC homeotic/homeobox/MADS, homeobox zipper, LIM domain, AP2
 CC and ERBS, zinc finger domains of type 2 Cys2His2, CCAAT box elements
 CC and MYB.
 XX
 SQ Sequence 348 BP; 106 A; 62 C; 87 G; 93 T; 0 other;
 Query Match 89.3%; Score 13.4; DB 21; Length 348;
 Best Local Similarity 93.3%; Pred. No. 8.3e+02;
 Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 AGCCATACATGAC 15
 ID AACC54150 standard; DNA; 428 BP.
 XX
 AC AACC54150;
 XX
 DT 18-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana DNA fragment SEQ ID NO: 76865.
 XX
 KW Hybridisation assay; genetic mapping; gene expression control;
 KW protein identification; signal transduction pathway;
 KW metabolic pathway; promoter; termination sequence; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 PN EP1033405-A2.
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0125788.
 PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 89.3%; Score 13.4; DB 21; Length 428;
Best Local Similarity 93.3%; Pred. No. 8.4e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
| | | | | | | | | | | | | | |
Db 419 AGCCATAACATAAAC 405

RESULT 9
ID AAA43836 standard; cDNA; 521 BP.
XX AAA43836;
XX 21-AUG-2000 (first entry)
DE Chicken secreted expressed sequence tag SEQ ID NO:411.
XX

KW Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;
KW expressed sequence tag; EST; probe; chemotactic; proliferative;
KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;
KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;
KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;
KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;
KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;
KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;
KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;
KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;
KW central nervous system disorder; Alzheimer's disease; stroke;
KW Parkinson's disease; Huntington's disease; coagulation disorder;
KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;
KW tumour; infection; depression; psoriasis; ss.

OS Gallus sp.
XX
XX WO200021991-A1.
PN
XX
XX 20-APR-2000.
PD
XX
PF 15-OCT-1999; 99WO-US24206.
XX
XX 15-OCT-1998; 98US-0104436.
PR

XX
PA (GEMY) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
PI Merberg D, Treacy M, Bowman MR;
XX
DR WPI; 2000-317938/27.
XX
PT Isolated polynucleotides, and encoded proteins, comprising secreted
PT expressed sequence tags (SESTs), useful for treating various disorders
PT such as autoimmune, infectious, and central nervous system disorders -
XX
PS Claim 1; Page 310-311; 803pp; English.

AAA43426 to AAA45925 represent specifically claimed secreted expressed
CC sequence tags (SESTs), isolated from human, mouse, chicken and rat
CC tissue sources. The SESTs can have a range of activities depending on
CC the tissues they were isolated from. The activities include:
CC chemotactic; proliferative; immunomodulatory; haematopoietic;
CC chemokinetic; analgesic; haemostatic; thrombolytic; antiinflammatory;
CC cyostatic; antibacterial; antifungal; antiviral; antidiabetic;
CC antiasthmatic; vulnery; antiulcer; osteopathic; neuroprotective;
CC nootropic; antiparkinsonian; antipsoriatic; cerebroprotective;
CC anticonvulsant; and antidepressant. The SESTs can be used for gene
CC therapy and in vaccines. The SESTs are useful as probes for the
CC identification and isolation of full-length cDNAs and genomic DNA
CC molecules which correspond to the SESTs. Proteins encoded by the SESTs
CC are useful in assays for determining biological activity and raising
CC antibodies. They may be useful for treatment of autoimmune disorders
CC (multiple sclerosis, insulin dependent diabetes), allergic conditions
CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,
CC osteoporosis, osteoarthritis, central nervous system disorders
CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation
CC disorders (haemophilia, thrombosis), inflammatory disorders (Crohn's
CC disease), tumours, bacterial, fungal or viral infections, depression and
CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given
XX in the exemplification of the present invention.

Sequence 521 BP; 177 A; 104 C; 125 G; 115 T; 0 other;

Query Match 89.3%; Score 13.4; DB 21; Length 521;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
| | | | | | | | | | | | | | |
Db 387 AGCCGTAACATGAAC 401

RESULT 10
ID ABA62712/C
XX ABA62712 standard; DNA; 556 BP.
XX
AC ABA62712;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #11017.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
XX WO200157277-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.

PT 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 1; SEQ ID NO 11017; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
AC |||||
DB 34 AGCCATAACATAAAC 20

RESULT 11
ABA30013/c
ID ABA30013 standard; DNA; 556 BP.
XX
AC ABA30013;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #8479 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human

PT hearts -
XX
PS Claim 1; SEQ ID NO 8479; 530pp; English.
XX

CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
AC |||||
DB 34 AGCCATAACATAAAC 20

RESULT 12
AAK11091/c
ID AAK11091 standard; DNA; 556 BP.
XX
AC AAK11091;
XX
DT 05-NOV-2001 (first entry)
XX

DE Human brain expressed single exon probe SEQ ID NO: 11082.

XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX

OS Homo sapiens.

PN WO200157275-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00667.

XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-483446/52.

XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX

PS Example 4; SEQ ID NO: 11082; 650pp + Sequence Listing; English.

XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system

CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.

XX Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
Db 34 AGCCATACATTAAC 20

RESULT 13
AAK36915/c
ID AAK36915 standard; DNA; 556 BP.

XX AAK36915;

DT 06-NOV-2001 (first entry)

DE Human bone marrow expressed single exon probe SEQ ID NO: 11472.

KW Human; bone marrow expressed exon; gene expression analysis; probe;
microarray; cancer; leukaemia; lymphoma; myeloma; ss.

OS Homo sapiens.

PN WO200157276-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00668.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488900/53.

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -

PS Example 4; SEQ ID NO: 11472; 658bp + Sequence Listing; English.

CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.

XX Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
Db 34 AGCCATACATTAAC 20

RESULT 14
AAI17764/c
ID AAI17764 standard; DNA; 556 BP.

XX AAI17764;

DT 12-OCT-2001 (first entry)

DE Probe #7697 for gene expression analysis in human cervical cell sample.

KW Probe; human; microarray; gene expression; cervical epithelial cell;
cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

PA (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 7697; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATACATGAAC 15
Db 34 AGCCATACATTAAC 20

RESULT 15

AAI42716/c
ID AAI42716 standard; DNA; 556 BP.

XX AAI42716;

DT 17-OCT-2001 (first entry)

DE Probe #11402 used to measure gene expression in human placenta sample.

KW Probe; microarray; human; placenta; antenatal diagnosis;

KW genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 11402; 654bp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 556 BP; 146 A; 105 C; 106 G; 199 T; 0 other;

Query Match 89.3%; Score 13.4; DB 22; Length 556;
Best Local Similarity 93.3%; Pred. No. 8.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
 |||||
Db 34 AGCCATAACATAAAC 20

Search completed: June 26, 2003, 03:11:49
Job time : 87 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccataacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	4	US-08-918-148-36 Sequence 36, Appl
2	13.4	89.3	1047	4	US-09-198-955A-7 Sequence 7, Appli
3	13.4	89.3	1047	4	US-09-694-531-7 Sequence 7, Appli
4	13.4	89.3	1687	1	US-08-656-984A-33 Sequence 33, Appl
5	13.4	89.3	3355	2	US-08-933-821-3 Sequence 3, Appli
6	13.4	89.3	3355	3	US-08-960-507-3 Sequence 3, Appli
7	13.4	89.3	3355	4	US-09-136-828-3 Sequence 3, Appli
8	13.4	89.3	3355	4	US-09-332-928A-3 Sequence 3, Appli
9	13.4	89.3	3355	4	US-09-136-801-3 Sequence 3, Appli
10	13.4	89.3	3355	4	US-09-332-929-3 Sequence 3, Appli
11	13	86.7	477	4	US-09-227-357-129 Sequence 129, App
12	13	86.7	3106	4	US-08-840-466A-21 Sequence 21, Appl
13	13	86.7	3106	4	US-09-696-188B-21 Sequence 21, Appl
14	13	86.7	3131	4	US-08-840-466A-20 Sequence 20, Appl
15	13	86.7	3131	4	US-09-696-188B-20 Sequence 20, Appl
16	13	86.7	43360	4	US-09-453-702B-206 Sequence 206, App
17	13	86.7	45325	4	US-09-453-702B-261 Sequence 261, App
18	12.4	82.7	66	1	US-08-009-265-50 Sequence 50, Appl
19	12.4	82.7	66	1	US-08-009-265-51 Sequence 51, Appl
20	12.4	82.7	213	1	US-08-009-265-29 Sequence 29, Appl
21	12.4	82.7	213	5	PCR-US94-06079-45 Sequence 45, Appl
22	12.4	82.7	323	4	US-08-638-931-64 Sequence 64, Appl
23	12.4	82.7	324	4	US-08-905-223-264 Sequence 264, App
24	12.4	82.7	431	4	US-08-887-534A-3 Sequence 3, Appli
25	12.4	82.7	485	4	US-09-177-325-4 Sequence 4, Appli
26	12.4	82.7	485	4	US-09-411-812A-4 Sequence 4, Appli
27	12.4	82.7	485	4	US-09-590-113-4 Sequence 4, Appli

C	28	12.4	82.7	639	4	US-09-328-111-180	Sequence 180, App
C	29	12.4	82.7	970	3	US-08-888-077A-28	Sequence 28, Appl
C	30	12.4	82.7	1278	2	US-08-909-965C-4	Sequence 4, Appli
C	31	12.4	82.7	1536	4	US-09-352-990-17	Sequence 17, Appli
C	32	12.4	82.7	3088	1	US-08-418-444A-1	Sequence 1, Appli
C	33	12.4	82.7	3937	3	US-08-586-165-8	Sequence 8, Appli
C	34	12.4	82.7	9207	3	US-08-388-353-800	Sequence 800, App
C	35	12.4	82.7	9207	3	US-08-488-551B-800	Sequence 800, App
C	36	12.4	82.7	9468	1	US-08-325-547-10	Sequence 10, Appl
C	37	12.4	82.7	22671	4	US-08-976-259-14	Sequence 14, Appl
C	38	12.4	82.7	32768	4	US-08-961-527-71	Sequence 71, Appl
C	39	12.4	82.7	36651	4	US-09-738-894A-3	Sequence 3, Appli
C	40	12.4	82.7	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C	41	12.4	82.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
C	42	12	80.0	441	4	US-08-887-534A-88	Sequence 88, Appl
C	43	12	80.0	558	4	US-09-134-001C-456	Sequence 456, App
C	44	12	80.0	4973	4	US-09-381-862-6	Sequence 6, Appli
C	45	12	80.0	11384	4	US-08-961-527-45	Sequence 45, Appl

ALIGNMENTS

RESULT 1
US-08-918-148-36
Sequence 36, Application US/08918148A
Patent No. 6342220
GENERAL INFORMATION:
APPLICANT: Adams, Camellia
APPLICANT: W.
APPLICANT: Carter, Paul J.
APPLICANT: Fendly, Brian M.
APPLICANT: Gurney, Austin L.
TITLE OF INVENTION: Agonist Antibodies
FILE REFERENCE: P0979
CURRENT APPLICATION NUMBER: US/08/918,148A
CURRENT FILING DATE: 1997-08-25
NUMBER OF SEQ ID NOS: 79
SEQ ID NO 36
LENGTH: 15
TYPE: DNA
ORGANISM: artificial
FEATURE:
NAME/KEY: 12D5scfv VH CDR1
LOCATION: 1-15
OTHER INFORMATION:
US-08-918-148-36

Query Match 100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
Db 1 AGCCATAACATGAAC 15

RESULT 2
US-09-198-955A-7/c
Sequence 7, Application US/09198955A
Patent No. 6187580
GENERAL INFORMATION:
APPLICANT: Andersen, Lene N.
APPLICANT: Schulein, Martin
APPLICANT: Lange, Niels E.
APPLICANT: Bjornvad, Mads E.
APPLICANT: Moller, Soren
APPLICANT: Glad, Sanne O. S.
APPLICANT: Kauppinen, Markus S.
APPLICANT: Schnorr, Kirk
APPLICANT: Kongsbak, Lars
TITLE OF INVENTION: NO. 6187580el Pectate Lyases
FILE REFERENCE: 5378.200-US

```

; CURRENT APPLICATION NUMBER: US/09/198,955A
; CURRENT FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-198-955A-7

```

```

Query Match      89.3%; Score 13.4; DB 4; Length 1047;
Best Local Similarity 93.3%; Pred. No. 79;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY      1 AGCCATAACATGAAC 15
        ||| ||| ||| |||
Db      95 AGCCATAACATTAAC 81

```

```

RESULT 3
US-09-694-531-7/c
; Sequence 7, Application US/09694531
; Patent No. 6368843
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. 6368843el Pectate lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/09/694,531
; PRIOR APPLICATION NUMBER: 09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-09-694-531-7

```

```

Query Match      89.3%; Score 13.4; DB 4; Length 1047;
Best Local Similarity 93.3%; Pred. No. 79;

```

```

Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 AGCCATAACATGAAC 15
        ||| ||| ||| |||
Db      95 AGCCATAACATTAAC 81

```

```

RESULT 4
US-08-656-984A-33/c
; Sequence 33, Application US/08656984A
; Patent No. 5753502
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; APPLICANT: Kilgannon, Patrick D.
; TITLE OF INVENTION: ICAM-4 Materials and Methods
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/656,984A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/827,689
; FILING DATE: 27-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/889,724
; FILING DATE: 26-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/894,061
; FILING DATE: 05-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/009,266
; FILING DATE: 22-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/102,852
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,295
; FILING DATE: 18-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/485,604
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: WILLIAMS, JR. JOSEPH A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33321
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-656-984A-33

```

```

Query Match      89.3%; Score 13.4; DB 1; Length 1687;
Best Local Similarity 93.3%; Pred. No. 85;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 AGCCATAACATGAAC 15
|||||
Db 450 AGCCATAACATGAAC 436

RESULT 5

US-08-933-821-3
; Sequence 3, Application US/08933821
; Patent No. 5972338
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/933,821
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3355 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-933-821-3

Query Match 89.3%; Score 13.4; DB 2; Length 3355;
Best Local Similarity 93.3%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
|||||
Db 815 AGCCGTAACATGAAC 829

RESULT 6

US-08-960-507-3
; Sequence 3, Application US/08960507
; Patent No. 6057435
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/960,507
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3355 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-08-960-507-3

Query Match 89.3%; Score 13.4; DB 3; Length 3355;
Best Local Similarity 93.3%; Pred. No. 93;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGCCATAACATGAAC 15
|||||
Db 815 AGCCGTAACATGAAC 829

RESULT 7

US-09-136-828-3
; Sequence 3, Application US/09136828
; Patent No. 6350450
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Tie Ligands
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Winpatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/136,828
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: P1130R1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-3216
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3355 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; US-09-136-828-3

Query Match 89.3%; Score 13.4; DB 4; Length 3355;
Best Local Similarity 93.3%; Pred. No. 93;

ATTORNEY/AGENT INFORMATION:
 NAME: Boone, Laurel S.
 REGISTRATION NUMBER: 43,505
 REFERENCE/DOCKET NUMBER: 04995.0029-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3106 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 21:
 US-08-840-466A-21

Query Match 86.7%; Score 13; DB 4; Length 3106;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATAACATGA 13
 |||||
 Db 60 AGCCATAACATGA 72

RESULT 13
 US-09-696-188B-21

Sequence 21, Application US/09696188B
 Patent No. 6406885

GENERAL INFORMATION:
 APPLICANT: Stewart, C. Neal
 McKee, Marian L.
 O'Brien, Alison D.
 Wachtel, Marian R.

TITLE OF INVENTION: Method Of Stimulating An Immune Response
 By Administration Of Host Organisms That Express Intimin
 Alone Or As A Fusion Protein With One Or More Other
 Antigens.

NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 Dunner, L.L.P.
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/696,188B
 FILING DATE: 26-Oct-2000
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/840,466
 FILING DATE: 1997-04-18
 ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laurel S.
 REGISTRATION NUMBER: 43,505
 REFERENCE/DOCKET NUMBER: 04995.0029-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3106 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 21:
 US-09-696-188B-21
 Query Match 86.7%; Score 13; DB 4; Length 3106;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATAACATGA 13
 |||||
 Db 60 AGCCATAACATGA 72

RESULT 14

US-08-840-466A-20
 Sequence 20, Application US/08840466A
 Patent No. 6261561

GENERAL INFORMATION:
 APPLICANT: Stewart, C. Neal
 McKee, Marian L.
 O'Brien, Alison D.
 Wachtel, Marian R.

TITLE OF INVENTION: Method Of Stimulating An Immune Response
 By Administration Of Host Organisms That Express Intimin
 Alone Or As A Fusion Protein With One Or More Other
 Antigens.

NUMBER OF SEQUENCES: 27
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 Dunner, L.L.P.
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/840,466A
 FILING DATE: 18-Apr-1997
 CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Boone, Laurel S.
 REGISTRATION NUMBER: 43,505
 REFERENCE/DOCKET NUMBER: 04995.0029-00000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3131 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
 SEQUENCE DESCRIPTION: SEQ ID NO: 20:
 US-08-840-466A-20

Query Match 86.7%; Score 13; DB 4; Length 3131;
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGCCATAACATGA 13
 |||||
 Db 197 AGCCATAACATGA 209

RESULT 15

US-09-696-188B-20
 Sequence 20, Application US/09696188B

```

: Patent No. 6406885
:
: GENERAL INFORMATION:
: APPLICANT: Stewart, C. Neal
:           McKee, Marian L.
:           O'Brien, Alison D.
:           Wachtel, Marian R.
:
: TITLE OF INVENTION: Method Of Stimulating An Immune Response
:                   By Administration Of Host Organisms That Express Intimin
:                   Alone Or As A Fusion Protein With One Or More Other
:                   Antigens.
:
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
:             Dunner, L.L.P.
:   STREET: 1300 I Street, N.W., Suite 700
:   CITY: Washington
:   STATE: D.C.
:   COUNTRY: USA
:   ZIP: 20005-3315
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/696,188B
:   FILING DATE: 26-Oct-2000
:   CLASSIFICATION: <Unknown>
:
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: 08/840,466
:   FILING DATE: 1997-04-18
:
: ATTORNEY/AGENT INFORMATION:
:   NAME: Boone, Laural S.
:   REGISTRATION NUMBER: 43,505
:   REFERENCE/DOCKET NUMBER: 04995.0029-00000
:
: TELECOMMUNICATION INFORMATION:
:   TELEPHONE: (202) 408-4000
:   TELEFAX: (202) 408-4400
:
: INFORMATION FOR SEQ ID NO: 20:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 3131 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: single
:     TOPOLOGY: linear
:   MOLECULE TYPE: DNA (genomic)
:   SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-696-188B-20

```

```

Query Match      86.7%; Score 13; DB 4; Length 3131;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 AGCCATACATGA 13
      |||
Db      197 AGCCATACATGA 209

```

Search completed: June 26, 2003, 04:52:38
 Job time : 31.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agcatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	13.4	89.3	556	10	US-09-864-761-8479
2	13.4	89.3	655	10	US-09-897-306-5
3	13.4	89.3	661	10	US-09-897-306-6
C 4	13.4	89.3	776	10	US-09-910-943-412
5	13.4	89.3	1047	12	US-10-072-152-7
6	13.4	89.3	1391	9	US-09-938-842A-2733
C 7	13.4	89.3	1517	9	US-09-892-877-80
C 8	13.4	89.3	1517	9	US-09-948-783-79
C 9	13.4	89.3	1518	9	US-09-892-877-115
C 10	13.4	89.3	1518	9	US-09-948-783-115
C 11	13.4	89.3	1687	9	US-10-025-524-33
12	13.4	89.3	2000	10	US-09-887-576-99
13	13.4	89.3	2181	10	US-09-897-306-2
14	13.4	89.3	2354	9	US-10-335-819-1
15	13.4	89.3	3355	9	US-10-028-072-277
16	13.4	89.3	3355	9	US-10-121-049-277
17	13.4	89.3	3355	9	US-10-123-904-277
18	13.4	89.3	3355	9	US-10-140-470-277
19	13.4	89.3	3355	9	US-10-175-746-277

20	13.4	89.3	3355	9	US-10-176-918-277	Sequence 277, App
21	13.4	89.3	3355	9	US-10-176-921-277	Sequence 277, App
22	13.4	89.3	3356	9	US-10-137-865-277	Sequence 277, App
23	13.4	89.3	3355	9	US-10-140-474-277	Sequence 277, App
24	13.4	89.3	3355	9	US-10-142-431-277	Sequence 277, App
25	13.4	89.3	3355	9	US-10-143-114-277	Sequence 277, App
26	13.4	89.3	3355	9	US-10-140-002-277	Sequence 277, App
27	13.4	89.3	3355	9	US-10-142-419-277	Sequence 277, App
28	13.4	89.3	3355	9	US-10-123-262-277	Sequence 277, App
29	13.4	89.3	3355	9	US-10-142-423-277	Sequence 277, App
30	13.4	89.3	3355	9	US-10-121-050-277	Sequence 277, App
31	13.4	89.3	3355	9	US-10-141-755-277	Sequence 277, App
32	13.4	89.3	3355	9	US-10-143-032-277	Sequence 277, App
33	13.4	89.3	3355	9	US-10-123-108-277	Sequence 277, App
34	13.4	89.3	3355	9	US-10-123-236-277	Sequence 277, App
35	13.4	89.3	3355	9	US-10-123-261-277	Sequence 277, App
36	13.4	89.3	3355	9	US-10-140-921-277	Sequence 277, App
37	13.4	89.3	3355	9	US-10-140-928-277	Sequence 277, App
38	13.4	89.3	3355	9	US-10-121-045-277	Sequence 277, App
39	13.4	89.3	3355	9	US-10-123-292-277	Sequence 277, App
40	13.4	89.3	3355	9	US-10-123-903-277	Sequence 277, App
41	13.4	89.3	3355	9	US-10-124-819-277	Sequence 277, App
42	13.4	89.3	3355	9	US-10-124-822-277	Sequence 277, App
43	13.4	89.3	3355	9	US-10-140-925-277	Sequence 277, App
44	13.4	89.3	3355	9	US-10-160-498-277	Sequence 277, App
45	13.4	89.3	3355	9	US-10-121-041-277	Sequence 277, App

ALIGNMENTS

RESULT 1
US-09-864-761-8479/c
; Sequence 8479, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 8479
; LENGTH: 556
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC018552.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
US-09-864-761-8479

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 556;
Best Local Similarity 93.3%; Pred. No. 7e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AGCCATAACATGAAC 15
    |||||
Db 34 AGCCATAACATGAAC 20

```

```

RESULT 2
US-09-897-306-5
; Sequence 5, Application US/09897306
; Patent No. US20020123054A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Patterson, Chandra
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN ANGIOPOIETIN
; FILE REFERENCE: PC-0048 CIP
; CURRENT APPLICATION NUMBER: US/09/897,306
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PERL Program
; SEQ ID NO 5
; LENGTH: 655
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020123054A1 SBDA07003F3
; NAME/KEY: unsure
; LOCATION: 532, 536, 555, 618, 630, 635, 641-642
; OTHER INFORMATION: a, t, c, g, or other
US-09-897-306-5

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 655;
Best Local Similarity 93.3%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AGCCATAACATGAAC 15
    |||||
Db 492 AGCCGTAACATGAAC 506

```

RESULT 3
US-09-897-306-6

```

; Sequence 6, Application US/09897306
; Patent No. US20020123054A1
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Gorgone, Gina A.
; APPLICANT: Patterson, Chandra
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN ANGIOPOIETIN
; FILE REFERENCE: PC-0048 CIP
; CURRENT APPLICATION NUMBER: US/09/897,306
; CURRENT FILING DATE: 2001-07-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PERL Program
; SEQ ID NO 6
; LENGTH: 661
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020123054A1 SBDA05881F1
; NAME/KEY: unsure
; LOCATION: 475, 506, 510, 545, 547, 554, 581, 599-600, 603, 615, 633, 638, 658-659
; OTHER INFORMATION: a, t, c, g, or other
US-09-897-306-6

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 661;
Best Local Similarity 93.3%; Pred. No. 7.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AGCCATAACATGAAC 15
    |||||
Db 204 AGCCGTAACATGAAC 218

```

```

RESULT 4
US-09-910-943-412/c
; Sequence 412, Application US/09910943
; Patent No. US20020081610A1
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Briavanlou, Ali
; APPLICANT: Altman, Curtis
; TITLE OF INVENTION: Assays and Materials for Embryonic Gene Expression
; FILE REFERENCE: 7529/1G1480S1
; CURRENT APPLICATION NUMBER: US/09/910,943
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 742
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 412
; LENGTH: 776
; TYPE: DNA
; ORGANISM: Xenopus laevis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(776)
; OTHER INFORMATION: n may be a or g or c or t/u
US-09-910-943-412

```

```

Query Match      89.3%; Score 13.4; DB 10; Length 776;
Best Local Similarity 93.3%; Pred. No. 7.3e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 AGCCATAACATGAAC 15
    |||||
Db 238 ATCCATAACATGAAC 224

```

RESULT 5
US-10-072-152-7/c
; Sequence 7, Application US/10072152
; Patent No. US20020142438A1
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schultein, Martin

```
; APPLICANT: Lange, Niels E.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Moller, Soren
; APPLICANT: Glad, Sanne O. S.
; APPLICANT: Kauppinen, Markus S.
; APPLICANT: Schnorr, Kirk
; APPLICANT: Kongsbak, Lars
; TITLE OF INVENTION: No. US20020142438A1e1 Pectate Lyases
; FILE REFERENCE: 5378.200-US
; CURRENT APPLICATION NUMBER: US/10/072,152
; CURRENT FILING DATE: 2002-02-07
; PRIOR APPLICATION NUMBER: US/09/198,955
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1343/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,249
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; PRIOR APPLICATION NUMBER: 09/073,684
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 09/184,217
; PRIOR FILING DATE: 1998-11-02
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Bacillus sp.
US-10-072-152-7
```

```
Query Match      89.3%; Score 13.4; DB 12; Length 1047;
Best Local Similarity 93.3%; Pred. No. 7.5e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATAACATGAAC 15
         |||||
Db       95 AGCCATAACATAAAC 81
```

```
RESULT 6
US-09-938-842A-2733
; Sequence 2733, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2733
; LENGTH: 1391
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2733
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1391;
Best Local Similarity 93.3%; Pred. No. 7.8e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATAACATGAAC 15
         |||||
Db       281 AGCCATAACATAAAC 295
```

```
RESULT 7
US-09-892-877-80/c
; Sequence 80, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 80
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1145)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-892-877-80
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1517;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATAACATGAAC 15
         |||||
Db       756 AGCCAGAACATGAAC 742
```

```
RESULT 8
US-09-948-783-79/c
; Sequence 79, Application US/09948783
; Publication No. US2003010051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
```

```
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 79
; LENGTH: 1517
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1145)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-79
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1517;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATACATGAAC 15
        ||||| ||||| |||||
Db       756 AGCCAGACATGAAC 742
```

```
RESULT 9
US-09-892-877-115/c
; Sequence 115, Application US/09892877
; Publication No. US20030077809A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P1
; CURRENT APPLICATION NUMBER: US/09/892,877
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/437,658
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-11-10
; NUMBER OF SEQ ID NOS: 461
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 115
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1146)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-892-877-115
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1518;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATACATGAAC 15
        ||||| ||||| |||||
Db       757 AGCCAGACATGAAC 743
```

```
RESULT 10
US-09-948-783-115/c
; Sequence 115, Application US/09948783
; Publication No. US20030100051A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 97 Human secreted proteins
; FILE REFERENCE: P2028P2
; CURRENT APPLICATION NUMBER: US/09/948,783
; CURRENT FILING DATE: 2001-09-10
```

```
; PRIOR APPLICATION NUMBER: 60/231,846
; PRIOR FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: 09/892,877
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: 09/437,658
; PRIOR FILING DATE: 1999-11-10
; PRIOR APPLICATION NUMBER: PCT/US99/09847
; PRIOR FILING DATE: 1999-05-06
; PRIOR APPLICATION NUMBER: 60/085,093
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,094
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,105
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,180
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085,927
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,906
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,924
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,922
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,921
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,923
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,925
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,928
; PRIOR FILING DATE: 1998-05-18
; PRIOR APPLICATION NUMBER: 60/085,920
; PRIOR FILING DATE: 1998-05-18
; NUMBER OF SEQ ID NOS: 465
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 115
; LENGTH: 1518
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1146)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-948-783-115
```

```
Query Match      89.3%; Score 13.4; DB 9; Length 1518;
Best Local Similarity 93.3%; Pred. No. 7.9e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY      1 AGCCATACATGAAC 15
        ||||| ||||| |||||
Db       757 AGCCAGACATGAAC 743
```

```
RESULT 11
US-10-025-524-33/c
; Sequence 33, Application US/10025524
; Publication No. US20030068659A1
; GENERAL INFORMATION:
; APPLICANT: Gallatin, W. Michael
; TITLE OF INVENTION: ICAM-4 Materials and Methods
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
;
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/025,524
;   FILING DATE: 18-Dec-2001
;   CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/827,689
;   FILING DATE: 27-JAN-1992
;   APPLICATION NUMBER: US 07/889,724
;   FILING DATE: 26-MAY-1992
;   APPLICATION NUMBER: US 07/894,061
;   FILING DATE: 05-JUN-1992
;   APPLICATION NUMBER: US 08/009,266
;   FILING DATE: 22-JAN-1993
;   APPLICATION NUMBER: US 08/102,852
;   FILING DATE: 05-AUG-1993
;   APPLICATION NUMBER: US 08/245,295
;   FILING DATE: 18-MAY-1994
;   APPLICATION NUMBER: US 08/485,604
;   FILING DATE: 07-JUN-1995
;
; ATTORNEY/AGENT INFORMATION:
;   NAME: WILLIAMS, JR. JOSEPH A.
;   REGISTRATION NUMBER: 38,659
;   REFERENCE/DOCKET NUMBER: 27866/33321
;
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-474-6300
;   TELEFAX: 312-474-0448
;   TELEX: 25-3856
;
; INFORMATION FOR SEQ ID NO: 33:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1687 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
;   SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-10-025-524-33

Query Match      89.3%; Score 13.4; DB 9; Length 1687;
Best Local Similarity 93.3%; Pred. No. 8e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAAC 15
        ||||| |||||
Db      450 AGCCATACCATGAAC 436

RESULT 12
US-09-887-576-99
; Sequence 99, Application US/09887576
; Patent No. US20020144047A1
;
; GENERAL INFORMATION:
;   APPLICANT: Budworth, P.
;   APPLICANT: Brown, D.
;   APPLICANT: Chang, H.
;   APPLICANT: Zhu, T.
;   APPLICANT: Han, B.
;   APPLICANT: Wang, X.
;   APPLICANT: Cooper, Bret
;   TITLE OF INVENTION: Promoters for regulation of plant expression
;   FILE REFERENCE: 1360.001US1
;   CURRENT APPLICATION NUMBER: US/09/887,576
;   CURRENT FILING DATE: 2001-06-25
;   PRIOR APPLICATION NUMBER: US 60/213,848
;   PRIOR FILING DATE: 2000-06-23
;   PRIOR APPLICATION NUMBER: US 60/214,087
;   PRIOR FILING DATE: 2000-06-23
;   PRIOR APPLICATION NUMBER: US 60/258,692
;   PRIOR FILING DATE: 2000-12-29
;   NUMBER OF SEQ ID NOS: 875
;   SOFTWARE: FastSeq for Windows Version 4.0
```

```
;
; SEQ ID NO 99
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-887-576-99

Query Match      89.3%; Score 13.4; DB 10; Length 2000;
Best Local Similarity 93.3%; Pred. No. 8.1e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAAC 15
        ||||| ||||| |||
Db      908 AGCCATAACATAAAC 922

RESULT 13
US-09-897-306-2
; Sequence 2, Application US/09897306
; Patent No. US20020123054A1
;
; GENERAL INFORMATION:
;   APPLICANT: Hillman, Jennifer L.
;   APPLICANT: Gorgone, Gina A.
;   APPLICANT: Patterson, Chandra
;   APPLICANT: Murry, Lynn E.
;   TITLE OF INVENTION: HUMAN ANGIOPOIETIN
;   FILE REFERENCE: PC-0048 CIP
;   CURRENT APPLICATION NUMBER: US/09/897,306
;   CURRENT FILING DATE: 2001-07-02
;   NUMBER OF SEQ ID NOS: 14
;   SOFTWARE: PERL Program
;
; SEQ ID NO 2
; LENGTH: 2181
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID NO. US20020123054A1 2365223CB1
US-09-897-306-2

Query Match      89.3%; Score 13.4; DB 10; Length 2181;
Best Local Similarity 93.3%; Pred. No. 8.2e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAAC 15
        |||| | ||||| |||||
Db      552 AGCCGTAACATGAAC 566

RESULT 14
US-10-335-819-1
; Sequence 1, Application US/10335819
; Publication No. US20030104573A1
;
; GENERAL INFORMATION:
;   APPLICANT: Shimkets, Richard A.
;   APPLICANT: Jeffers, Michael
;   TITLE OF INVENTION: Nucleotide Sequences and Amino Acid Sequences of
;   FILE REFERENCE: 15966-517 CIP1
;   CURRENT APPLICATION NUMBER: US/10/335,819
;   CURRENT FILING DATE: 2003-01-02
;   PRIOR APPLICATION NUMBER: US/09/658,644
;   PRIOR FILING DATE: 2001-09-09
;   PRIOR APPLICATION NUMBER: U.S.S.N. 09/150,684
;   PRIOR FILING DATE: 1998-09-10
;   NUMBER OF SEQ ID NOS: 8
;   SOFTWARE: Patentin Ver. 2.1
;   SEQ ID NO 1
; LENGTH: 2354
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (352)..(1824)
```


; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082999
; PRIOR FILING DATE: 1998-04-24
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/083545
; PRIOR FILING DATE: 1998-04-29
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084627
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/084637
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/085149
; PRIOR FILING DATE: 1998-05-12
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085338
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085339
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085697
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/085704
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086414
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/086430
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088730
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088741
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088810
; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088858
; PRIOR FILING DATE: 19/98-06-11
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089599
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089907
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/089947
; PRIOR FILING DATE: 1998-06-19
; PRIOR APPLICATION NUMBER: 60/090349
; PRIOR FILING DATE: 1998-06-23
; PRIOR APPLICATION NUMBER: 60/090429
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090538
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07

Db | | | | | | | | | |
815 AGCCGTACATGAAC 829

Search completed: June 26, 2003, 04:57:26
Job time : 66.2 secs

Query Match 89.3%; Score 13.4; DB 9; Length 3355;
Best Local Similarity 93.3%; Pred. No. 8.6e+02;
Matches 14; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AGCCATACATGAAC 15

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-36
Perfect score: 15
Sequence: 1 agccatacatgaac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_trod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	109	13	BJ526500	BJ526500 BJ526500
2	15	100.0	396	17	AQ931789	AQ931789 RPCI-23-2
3	15	100.0	473	17	AQ664590	AQ664590 HS_5220_B
4	15	100.0	539	17	AQ312321	AQ312321 RPCI11-11
5	15	100.0	609	17	AZ620907	AZ620907 1M0453M18
6	15	100.0	637	17	BH616021	BH616021 BMBAC305H

C	7	15	100.0	657	17	AZ572103	AZ572103 303PVD05
C	8	15	100.0	724	13	BM038276	BM038276 U006D06 O
C	9	15	100.0	754	17	AG123454	AG123454 Pan trogl
C	10	15	100.0	891	17	CNS04EDW	AL286925 Tetraodon
C	11	14	93.3	288	10	BE326934	BE326934 hr67g09.x
C	12	14	93.3	350	9	AA953965	AA953965 0089g06.s
C	13	14	93.3	381	10	AW433346	AW433346 sh55h12.Y
C	14	14	93.3	384	14	BQ048783	BQ048783 952024A09
C	15	14	93.3	403	17	BH863054	BH863054 SALK_0930
C	16	14	93.3	404	12	BG413078	BG413078 ETESTed13
C	17	14	93.3	414	17	BH461356	BH461356 BOHOY93TR
C	18	14	93.3	424	10	AV418590	AV418590 AV418590
C	19	14	93.3	428	12	BG378431	BG378431 UT-R-CU0-
C	20	14	93.3	465	17	DR10G4T	AL743891 Danilo rer
C	21	14	93.3	468	10	AW928466	AW928466 EST337254
C	22	14	93.3	468	17	AO885515	AO885515 HS_5524_A
C	23	14	93.3	473	12	BG346905	BG346905 dad19c03.
C	24	14	93.3	485	17	AO760183	AO760183 HS_2026_A
C	25	14	93.3	485	17	GG0000346	AL605866 Gallus ga
C	26	14	93.3	488	17	GG0000497	AL606017 Gallus ga
C	27	14	93.3	494	12	BG348831	BG348831 dad19c03.
C	28	14	93.3	503	17	AO627757	AO627757 CITBI-E1-
C	29	14	93.3	511	10	AW638003	AW638003 b165a09.w
C	30	14	93.3	522	10	AW928608	AW928608 EST337396
C	31	14	93.3	529	17	BH037756	BH037756 RPCI-24-3
C	32	14	93.3	536	10	AW719266	AW719266 LJNEST1h8
C	33	14	93.3	550	13	BI893081	BI893081 sai61b01.
C	34	14	93.3	559	17	FR0026576	AL019407 F.rubripe
C	35	14	93.3	564	10	AW720673	AW720673 LJNEST4b2
C	36	14	93.3	566	10	BE659030	BE659030 GM700008A
C	37	14	93.3	568	17	B00976	B00976 CSRL-122g1-
C	38	14	93.3	577	13	BI416305	BI416305 LJNEST4b2
C	39	14	93.3	580	13	BM023966	BM023966 fu67g11.Y
C	40	14	93.3	592	17	FR0030167	AL026536 Fugu rubr
C	41	14	93.3	599	10	AV720297	AV720297 AV720297
C	42	14	93.3	602	17	AO978044	AO978044 RPCI-23-3
C	43	14	93.3	610	17	GG0000457	AL605977 Gallus ga
C	44	14	93.3	611	14	BQ852431	BQ852431 OGB18A04.
C	45	14	93.3	613	13	BJ037212	BJ037212 BJ037212

ALIGNMENTS

RESULT 1
BJ526500
LOCUS
DEFINITION BJ526500 109 bp mRNA linear EST 09-AUG-2002
mRNA sequence.
ACCESSION BJ526500
VERSION BJ526500
KEYWORDS BJ526500.1 GI:22185312
SOURCE EST.
ORGANISM Japanese medaka.

ORGANISM

Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
Belontiiformes; Adrianichthyidae; Oryziinae; Oryzias.

REFERENCE 1 (bases 1 to 109)
AUTHORS Kohara,Y., Shin-i,T., Kimura,T., Narita,T., Jindo,T. and Takeda,H.
TITLE Medaka EST Project in Takeda's lab
JOURNAL Unpublished (2001)
COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855

FEATURES
Source location/Qualifiers
1..109
/organism="Oryzias latipes"
/strain="Hd-tr"

/db_xref="taxon:8090"
/clone="MF01SSB020H21"
/clone_lib="MF01SSB cDNA"
/sex="mixture of female and male"
/tissue_type="whole embryo"
/dev_stage="segmentation stage 20 - 25"
31 a 33 c 22 g 22 t 1 others

Query Match 100.0%; Score 15; DB 13; Length 109;
Best Local Similarity 100.0%; Pred. No. 8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
16 AGCCATAACATGAAC 30

RESULT 2
LOCUS AQ931789 396 bp DNA linear GSS 21-DEC-1999
DEFINITION RPCI-23-269112.TV RPCI-23 Mus musculus genomic clone RPCI-23-269112
ACCESSION AQ931789
VERSION AQ931789.1 GI:6620803
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 396)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-269112.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 269 row: I column: 12
Seq primer: T7
Class: BAC ends.

FEATURES
source location/Qualifiers
1. .396
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-269112"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 102 a 105 c 75 g 114 t
ORIGIN
Query Match 100.0%; Score 15; DB 17; Length 396;

Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
5 AGCCATAACATGAAC 19

RESULT 3
LOCUS AQ664590/c 473 bp DNA linear GSS 23-JUN-1999
DEFINITION HS_5220_B1_H03-SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=796 Col=5 Row=P, DNA sequence.
ACCESSION AQ664590
VERSION AQ664590.1 GI:5172358
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 473)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
Hood,L.
TITLE Sequence-tagged connectors: A sequence approach to mapping and
scanning the human genome
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
COMMENT 99380589
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 796 row: P column: 5
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 473.

FEATURES
source location/Qualifiers
1. .473
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=796 Col=5 Row=P"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
Male blood DNA was isolated from one randomly chosen donor
and partially digested with a combination of EcoRI and
EcoRI Methylase. Size selected DNA was cloned into the
pBACe3.6 vector at EcoRI sites"

BASE COUNT 134 a 89 c 95 g 155 t
ORIGIN
Query Match 100.0%; Score 15; DB 17; Length 473;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
Db 354 AGCCATAACATGAAC 340

RESULT 4
LOCUS AQ312321/c 539 bp DNA linear GSS 04-MAY-1999
DEFINITION RPCI11-112M12.TJ RPCI-11 Homo sapiens genomic clone RPCI-11-112M12,

DNA sequence.
ACCESSION AQ312321
VERSION AQ312321.1 GI:4043985
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 539)
Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and Venter,J.C. Use of human BAC End Sequences for Sequence-Ready Map Building
JOURNAL Unpublished (1998)
TITLE Other_GSSs: RPC11-112M12.TV
COMMENT Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: SP6
Class: BAC ends.
FEATURES
source
1..539
/organism="Homo sapiens"
/db_xref="GDB:7542923"
/db_xref="taxon:9606"
/clone="RPCI-11-112M12"
/clone_lib="RPCI-11"
/sex="Male"
/cell_type="Lymphocytes"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; RPC11 Human Male BAC Library"
BASE COUNT 163 a 97 c 106 g 173 t
ORIGIN
Query Match 100.0%; Score 15; DB 17; Length 539;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCATACATGAAC 15
|||||
Db 357 AGCCATACATGAAC 343
RESULT 5
AZ620907 609 bp DNA linear GSS 13-DEC-2000
LOCUS 1M0453M18R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0453M18 R, DNA sequence.
ACCESSION AZ620907
VERSION AZ620907.1 GI:11743097
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 609)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C., Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A. and Wright,D., Weiss,R.
TITLE Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT 84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0453 row: M column: 18
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 609.
FEATURES
source
1..609
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0453M18"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of PWD42 (g1473211419b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT 192 a 126 c 113 g 178 t
ORIGIN
Query Match 100.0%; Score 15; DB 17; Length 609;
Best Local Similarity 100.0%; Pred. No. 1.5e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCCATACATGAAC 15
|||||
Db 144 AGCCATACATGAAC 158
RESULT 6
BH616021/c 637 bp DNA linear GSS 28-JAN-2002
LOCUS BMBAC305H05SP6_PSU Brugia malayi Genomic Bac Library 3 Brugia
DEFINITION malayi genomic, DNA sequence.
ACCESSION BH616021
VERSION BH616021.1 GI:18380709
KEYWORDS GSS.
SOURCE Brugia malayi.
ORGANISM Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Brugia.
REFERENCE 1 (bases 1 to 637)
Whitton,C., Daub,J., Ware,J., Quail,M., Hall,N., Barrell,B., Foster,J., Guillano,D., Slatko,B. and Blaxter,M.
TITLE Genome survey sequences from the human parasitic nematode Brugia malayi
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh

Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
Sequenced from the Brugia malayi BAC library constructed by Claire Whitton and Dr Mike Quail. The sequence was generated by The Pathogen Sequencing Unit, The Sanger Institute, Cambridge, UK in collaboration with Mark Blaxter, ICAPB, University of Edinburgh, Edinburgh, UK.
Seq primer: SP6 (ATTAGGTGACACTATAG)
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1. 637
/organism="Brugia malayi"
/strain="TRS"
/db_xref="taxon:6279"
/clone_lib="Brugia malayi Genomic Bac Library 3"
/sex="Mixed (male and female)"
/tissue_type="whole parasite"
/dev_stage="microfilaria (L1)"
/note="Vector: pBACE3.6; Site_1: BamH I; Brugia malayi genomic DNA was partially cleaved with Sau3A I and size fractionated. 7,392 clones were generated with mean insert size ~48 kbp. The library was constructed by Claire Whitton, Blaxter Nematode Genetics Lab, University of Edinburgh, UK, and Dr Mike Quail, The Pathogen Sequencing Unit, The Sanger Centre, Cambridge, UK."

BASE COUNT
ORIGIN
196 a 109 c 120 g 212 t

Query Match
Best Local Similarity 100.0%; Score 15; DB 17; Length 637;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
Db 489 AGCCATAACATGAAC 475

RESULT 7
AZ572103/c 657 bp DNA linear GSS 15-MAY-2001
LOCUS 303PVD05 PV MBN #30 Plasmodium vivax genomic 3', DNA sequence.
ACCESSION AZ572103
VERSION AZ572103.1 GI:13985046
KEYWORDS GSS.
SOURCE malaria parasite P. vivax.
ORGANISM Plasmodium vivax
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 657)
TITLE Carlton,J.M.-R. and Dame,J.B.
JOURNAL The Plasmodium vivax and P. berghei gene sequence tag projects
COMMENT Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

FEATURES
source
Location/Qualifiers
1. 657
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"
/db_xref="taxon:5855"
/clone_lib="PV MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/note="Vector: pBluescript SK(+) vector DNA, phagemid

excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V;
Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmidpur filter, followed by passage through a column of pre-wet Whatman CF11 powder (1:2 ratio volume of blood to CF11), and finally centrifuged through a 50% percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 500C as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA Polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

BASE COUNT
ORIGIN
147 a 146 c 148 g 213 t 3 others

Query Match
Best Local Similarity 100.0%; Score 15; DB 17; Length 657;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
Db 294 AGCCATAACATGAAC 280

RESULT 8
BM038276/c 724 bp mRNA linear EST 06-NOV-2001
LOCUS U006D06 Oryza sativa mature leaf library induced by M.grisea Oryza
DEFINITION sativa cDNA clone U006D06, mRNA sequence.
ACCESSION BM038276
VERSION BM038276.1 GI:16753897
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 724)
AUTHORS Dong,H.T., Li,D.B., Zhuang,X.F., Dai,C.G., Sun,L.X., Pei,Y.X., Wu ,H.F., Jiang,Y.X., Yu,F.C., Gao,Q.K. and Lou,Y.C.
TITLE A Gene Expression Screen in Oryza sativa
JOURNAL Unpublished (2001)
COMMENT Contact: Dong HT
Laboratory of Functional Genetics
Bio-technology Institute of Zhejiang University
Kaixuan Road 268#, Hangzhou, Zhejiang, P.R.China
Tel: 0086-571-86892051
Fax: 0086-571-86961525
Email: htdong@zjuem.zju.edu.cn
Seq primer: M13 forward primer.

FEATURES
source
Location/Qualifiers
1. 724
/organism="Oryza sativa"
/db_xref="taxon:4530"
/clone="U006D06"
/clone_lib="Oryza sativa mature leaf library induced by M.grisea"
/tissue_type="leaf"
/dev_stage="Mature stage"
/note="Vector: pSport2"

BASE COUNT
ORIGIN
149 a 181 c 185 g 206 t 3 others

Query Match
Best Local Similarity 100.0%; Score 15; DB 13; Length 724;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
|||||
Db 53 AGCCATAACATGAAC 39

RESULT 9
AG123454/c
LOCUS AG123454 754 bp DNA linear GSS 04-NOV-2001
DEFINITION Pan troglodytes DNA, clone: PTB-133B19.R, genomic survey sequence.
ACCESSION AG123454
VERSION AG123454.1 GI:16652619
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-133B19.R.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE BAC end sequences of Library PTB
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 754)
AUTHORS Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC): 1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:chimbesc@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT Clones are derived from the chimpanzee BAC library PTB This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
PRIMERS

COMMENT Sequencing: M13Rev
LIBRARY Vector : PKS145
R.Site 1 : SacI
R.Site 2 : SacI.
FEATURES Location/Qualifiers
source 1..754
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-133B19.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"

BASE COUNT .258 a 124 c 127 g 245 t
ORIGIN

Query Match 100.0%; Score 15; DB 17; Length 754;
Best Local Similarity 100.0%; Pred. No. 1.6e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
|||||
Db 585 AGCCATAACATGAAC 571

RESULT 10
CNS04EDM/c
LOCUS CNS04EDM 891 bp DNA linear GSS 21-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone 104K08 of library G from Tetraodon nigroviridis, genomic survey sequence.
ACCESSION AL286925
VERSION AL286925.1 GI:8025396
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

FEATURES
source
1..891
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="104K08"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG104BF04SP1-end : PUC-Ori"

BASE COUNT 164 a 252 c 247 g 211 t 17 others
ORIGIN

Query Match 100.0%; Score 15; DB 17; Length 891;
Best Local Similarity 100.0%; Pred. No. 1.7e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCATAACATGAAC 15
|||||
Db 490 AGCCATAACATGAAC 476

RESULT 11
BE326934
LOCUS BE326934 288 bp mRNA linear EST 14-JUL-2000
DEFINITION hrf67g09.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3133600 3', mRNA sequence.
ACCESSION BE326934
VERSION BE326934
KEYWORDS BE326934.1 GI:9200710
SOURCE EST.
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 288)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaabs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLML, send email to: info@image.llnl.gov
Seq primer: -40UP from gibco.
FEATURES Location/Qualifiers

```

source
1. .288
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3133600"
/clone_lib="NCI_CGAP_Kid1"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT      102 a      70 c      30 g      86 t
ORIGIN
Query Match      93.3%; Score 14; DB 10; Length 288;
Best Local Similarity 100.0%; Pred. No. 3.8e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAA 14
        |||||||
Db      191 AGCCATAACATGAA 204

RESULT 12
AA953965      350 bp      mRNA      linear      EST 07-JUL-1998
LOCUS      co088906.s1 NCI_CGAP_Kid5 Homo sapiens CDNA clone IMAGE:1573306 3,
DEFINITION      similar to gb:X56740 RAS-RELATED PROTEIN RAB-11 (HUMAN);, mRNA
sequence.
AA953965
AA953965.1 GI:3116883
EST.
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 350)
AUTHORS      NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL      Unpublished (1997)
COMMENT      Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/HLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 952 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 147.
FEATURES
source
1. .350
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1573306"
/clone_lib="NCI_CGAP_Kid5"
/tissue_type="2 pooled tumors (clear cell type)"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAAGAATTGGCGCGCATATTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
```

```

and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      127 a      60 c      72 g      91 t
ORIGIN
Query Match      93.3%; Score 14; DB 9; Length 350;
Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGCCATAACATGAA 14
        |||||||
Db      35 AGCCATAACATGAA 48

RESULT 13
AW433346/c      381 bp      mRNA      linear      EST 03-DEC-2001
LOCUS      sh5sh12.Y1 Gm-c1015 glycine max CDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION      Gm-c1015-3744 5' similar to TR:P96167 P96167 PUTATIVE ALDOLASE. ;,
mRNA sequence.
AW433346
AW433346.1 GI:6964653
EST.
KEYWORDS      soybean.
SOURCE      Glycine max
ORGANISM      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE      1 (bases 1 to 381)
AUTHORS      Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
TITLE      Public Soybean EST Project
JOURNAL      Unpublished (1999)
COMMENT      Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
Insert Length: 730 Std Error: 0.00
Seq primer: -40RP from Gibco
High quality sequence stop: 336.
FEATURES
source
1. .381
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1015-3744"
/clone_lib="Gm-c1015"
/tissue_type="Mature flowers, field grown plants"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This CDNA library was constructed from mRNA isolated
from mature flowers of field grown plants. The CDNA
library was prepared using the Stratagene pBluescript II
XR cDNA library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adaptors
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
```

	Erpelding."
BASE COUNT	109 a 67 c 98 g 107 t
ORIGIN	

Query Match	93.38;	Score 14;	DB 10;	Length 381;
Best Local Similarity	100.0%;	Pred. No. 4.2e+03;		
Matches 14; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	2	GCCATACATGAAC	15
Db	314	GCCATACATGAAC	301

RESULT 14
BQ048783

LOCUS	BQ048783	384 bp	mRNA	linear	EST 29-MAR-2002
DEFINITION	952024A09.Y2 952 - BMS tissue from walbot lab (reduced rRNA) Zea mays cDNA, mRNA sequence.				

ACCESSION	BQ048783
VERSION	BQ048783.1
KEYWORDS	GI:19822759 EST.

SOURCE	Zea mays.
ORGANISM	Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 384)

TITLE
Maize ESTs from various cDNA libraries sequenced at Stanford
University

JOURNAL
COMMENT

unpublished (1999)
Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 952024 row: A column: 09.

FEATURES.	Location/Qualifiers
source	1. .384

```

/organism="Zea mays"
/cultivar="BMS (Black Mexican Sweet)"
/db_xref="taxon:4577"
/clone_lib="952 - BMS tissue from Walbot Lab (reduced rRNA)
"
```

```
/tissue_type="suspension culture"  
/dev_stage="mixed logarithmic and stationary growth  
phases"
```

/note="Vector: pUC19; Site_1: EcoRI; Site_2: EcoRI; The library was prepared by George Rudenko using poly (A) selected RNA and Universal Riboclone cDNA Synthesis System (Promega). cDNA was synthesized using both random and oligo(dT) primers in separate reactions and equipped with EcoRI adaptors. Library was size-fractionated on agarose gels (for insert size >400bp) and non-directionally cloned into EcoRI-digested pUC19 vector. Blue/white selection on carbenicillin-containing plates was used to recover positive clones."

BASE COUNT	91 a	102 c	62 g	129 t
ORIGIN				

```

Query Match          93.3%;  Score 14;  DB 14;  Length 384;
Best Local Similarity 100.0%;  Pred. No. 4.2e+03;
Matches 14;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY      1 AGCCATACATGAA 14
      .  |||||
Db      358 AGCCATACATGAA 371

```

RESULT 15

BH863054/c	403 bp	DNA	linear	GSS 05-AUG-2002
LOCUS	BH863054			
DEFINITION	Arabidopsis thaliana TDNA insertion lines Arabidopsis			

ACCESSION	BH863054
VERSION	BH863054.1
KEYWORDS	GI:22098383 GSS.

ORGANISM

REFERENCE
AUTHORS
Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab
1 (bases 1 to 403)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 403)

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL
COMMENT

Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of
TDNA.
Class: TDNA tagged.

FEATURES
source

```
1. .403
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_093049"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
```

BASE COUNT	92 a	102 c	84 g	125 t
ORIGIN				

```
Query Match          93.3%; Score 14; DB 17; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.3e+03;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

QY      2  GCCATACATGAAC  15
          |||||
Db      245 GCCATACATGAAC  232

```

Search completed: June 26, 2003, 04:50:18
Job time : 753.2 secs

Search completed: June 26, 2003, 04:50:18
Job time : 753.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:01:10 ; Search time 425.6 Seconds
(without alignments)
1025.710 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agttactactggagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	6	AR183495
2	15	100.0	209	9	HS1GVHG41
3	15	100.0	218	9	HSJ227742
4	15	100.0	219	9	HSU86367
5	15	100.0	220	9	HSJ227729
6	15	100.0	221	9	HSVH4GL7
7	15	100.0	225	9	S75891
8	15	100.0	229	9	HS347134
9	15	100.0	230	9	HSVH4GL11
10	15	100.0	234	9	HSVH4GL10
11	15	100.0	234	9	HSVH4GL9
12	15	100.0	235	9	HSVH4GL1
13	15	100.0	236	9	HS1GH2B1
14	15	100.0	237	9	HSBDP71RB
15	15	100.0	238	9	HS347131
16	15	100.0	238	9	HS347367
17	15	100.0	238	9	HS347368
18	15	100.0	240	9	HS347352
19	15	100.0	240	9	HS347122
20	15	100.0	241	9	HS347352
21	15	100.0	244	9	HS347166
22	15	100.0	248	9	S74500
23	15	100.0	249	9	AY013307
24	15	100.0	251	9	HS347242
25	15	100.0	253	9	HS298472
26	15	100.0	253	9	HS414778
27	15	100.0	256	9	HS402538
28	15	100.0	256	9	HSVIGD2
29	15	100.0	258	9	HS402422
30	15	100.0	261	9	HS402380
31	15	100.0	262	9	HS414770
32	15	100.0	262	9	S74642
33	15	100.0	264	9	HSEM116V4
34	15	100.0	265	9	HS402546
35	15	100.0	267	9	HSVIGD4
36	15	100.0	267	9	HSVIGD6
37	15	100.0	268	9	HS347149
38	15	100.0	268	9	HS402470
39	15	100.0	270	9	HS1G18V4
40	15	100.0	270	9	HS280554
41	15	100.0	272	9	HS402535
42	15	100.0	273	9	HS1G31V4
43	15	100.0	276	9	AF077488
44	15	100.0	276	9	HSVCB9
45	15	100.0	276	9	HS280734

ALIGNMENTS

RESULT 1
AR183495
LOCUS AR183495 15 bp DNA
DEFINITION Sequence 42 from patent US 6342220.
ACCESSION AR183495
VERSION AR183495.1 GI:20227464
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 15)
AUTHORS Adams,C.W., Carter,P.J., Fendly,B.M. and Gurney,A.L.
TITLE Agonist antibodies
JOURNAL Patent: US 6342220-A 42 29-JAN-2002;
FEATURES Location/Qualifiers

[illegible]

CDS	/organism="Homo sapiens" /db_xref="taxon:9606" /rearranged <1..>218 /codon_start=2 /product="immunoglobulin heavy chain variable region" /protein_id="CA12720.1" /db_xref="GI:2951633" /translation="GGSVSSYYWSWIRPEPGKLEWIGYIFYTGSTNYNPSLSKSRVTI SLDMSDQFSLKLSVTAADTAVYYCAR" <1..>218 /product="immunoglobulin heavy chain variable region"									
BASE COUNT	48	a	61	c	60	g	49	t		
ORIGIN										
Query Match	100.0%; Score 15; DB 9; Length 218;									
Best Local Similarity	100.0%; Pred. NO. 6.1e+02;									
Matches	15;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	1	AGTTACTACTGGAGC	15							
Db	17	AGTTACTACTGGAGC	31							
RESULT 4										
HSU86367										
LOCUS	HSU86367	219	bp	DNA	linear	PRI 15-FEB-1997				
DEFINITION	Human rearranged IGH variable region gene, partial cds.									
ACCESSION	U86367									
VERSION	U86367.1	GI:1841799								
KEYWORDS										
SOURCE	Homo sapiens.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.									
AUTHORS	1 (bases 1 to 219)									
TITLE	Tiersens,A. and Delabie,J.									
JOURNAL	Marginal zone lymphoma 8-VH4									
REFERENCE	2 (bases 1 to 219)									
AUTHORS	Tiersens,A. and Delabie,J.									
TITLE	Direct Submission									
JOURNAL	Submitted (22-JAN-1997) Pathology, University of Leuven, Minderbroedersstraat 12, Leuven B-3000, Belgium									
FEATURES	Location/Qualifiers									
source	1..219									
	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	/rearranged									
	/note="marginal zone lymphoma 8-VH4"									
	<1..>219									
	/codon_start=1									
	/product="IGH variable region"									
	/protein_id="AAB47529.1"									
	/db_xref="GI:1841800"									
	/translation="SGGSISYYWSMIROSPGKLEWIGYIYSSGSTNYNPSLSKRYT ISVDTSKNQFSLKLSVTAADTAVYYCAR"									
BASE COUNT	53	a	54	c	56	g	56	t		
ORIGIN										
Query Match	100.0%; Score 15; DB 9; Length 219;									
Best Local Similarity	100.0%; Pred. NO. 6.1e+02;									
Matches	15;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;	
QY	1	AGTTACTACTGGAGC	15							
Db	19	AGTTACTACTGGAGC	33							
RESULT 5										
LOCUS	HSJ227729	220	bp	DNA	linear	PRI 10-MAR-1998				
DEFINITION	Homo sapiens DNA for rearranged immunoglobulin heavy chain gene,									

2-11VH4.
AJ227729 GI:2951609
heavy chain; immunoglobulin; variable region.
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 220)
Delabie,J.
Direct Submission
Submitted (03-MAR-1998) Delabie J., Pathology, University of
Leuven, Minderbroedersstraat 12, B-3000, BELGIUM
2 (bases 1 to 220)
Delabie,J. and Tiersens,A.
Evidence for clonal expansion and somatic hypermutations of the
marginal zone B cells in the lymph node and spleen
Unpublished
Location/Qualifiers
1..220
/organism="Homo sapiens"
/db_xref="taxon:9606"
/rearranged
<1..>220
/codon_start=3
/product="immunoglobulin heavy chain variable region"
/protein_id="CAA12708.1"
/db_xref="GI:2951610"
/translation="GGSSISYWSWIRPPGKLEWIGIYYSGSTNYNPSLKSQVTI
SVDTSKNQFSLKLSSTVTAADTAVYYCAR"
<1..>220
/product="immunoglobulin heavy chain variable region"
V_region
BASE COUNT 54 a 60 c 60 g 46 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 220;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
|||||
Db 18 AGTTACTACTGGAGC 32
RESULT 6
HSVH4GL7 221 bp DNA linear PRI 27-JUN-1996
LOCUS H.sapiens germline immunoglobulin heavy chain VH gene (clone
VH4-GL7).
Z75359
ACCESSION Z75359.1 GI:1418960
VERSION immunoglobulin; immunoglobulin heavy chain; immunoglobulin
KEYWORDS superfamily; immunoglobulin variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 221)
Voswinkel,J., Truemper,L., Carbon,G., Hopf,T., Pfreundschuh,M. and
Gause,A.
Evidence for a selected humoral immune response encoded by VH4
family genes in the synovial membrane of a patient with rheumatoid
arthritis
Unpublished
2 (bases 1 to 221)
Voswinkel,J.
Direct Submission
Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland
Medical School, D-66421 Homburg/Saar, GERMANY
Location/Qualifiers
1..221
/organism="Homo sapiens"
/isolate="patient 8 with rheumatoid arthritis"

/db_xref="taxon:9606"
/chromosome="14"
/clone="VH4-GL7"
/cell_type="B lymphocyte"
/tissue_type="synovial membrane"
/clone_lib="VH4-GL"
/germline
V_segment 1..221
BASE COUNT 51 a 60 c 59 g 47 t 4 others
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 221;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
|||||
Db 19 AGTTACTACTGGAGC 33
RESULT 7
S75891 225 bp mRNA linear PRI 27-JUL-1995
LOCUS IGH-immunoglobulin heavy chain variable region [clone M1] (human,
DEFINITION multiple myeloma patient, mRNA Partial, 225 nt).
ACCESSION S75891
VERSION S75891.1 GI:913791
KEYWORDS S75891.1 GI:913791
SOURCE Homo sapiens multiple myeloma patient.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 225)
Baker,B.W., Deane,M., Gilleece,M.H., Johnston,D., Scarffe,J.H. and
Norton,J.D.
Distinctive features of immunoglobulin heavy chain variable region
gene rearrangement in multiple myeloma
Leuk. Lymphoma 14 (3-4), 291-301 (1994)
95038365
PUBMED 7950918
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI g1bbsq 161439] from the original journal article.
FEATURES
source
1..225
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..225
gene
1..225
/partial
/gene="IGH"
/note="immunoglobulin heavy chain variable region"
1..225
/partial
/gene="IGH"
/note="immunoglobulin heavy chain variable region"
1..225
/partial
/gene="IGH"
/note="immunoglobulin heavy chain variable region"
/codon_start=1
/protein_id="AAD14200.1"
/db_xref="GI:4261900"
/translation="AVSGYSISYWSWIRQSPKGLDFIAYFHNSRGTTYNPSLKS
ATISVDTSKNQFSLRLSSVTADTAVYYCAR"
BASE COUNT 48 a 67 c 56 g 54 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 225;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGGAGC 15
|||||
Db 25 AGTTACTACTGGAGC 39
RESULT 8
HSA347134

LOCUS HSA347134 229 bp DNA linear PRI 26-MAR-2002
DEFINITION Homo sapiens partial IGWH4-61 gene for immunoglobulin heavy chain variable region, isolate case2-cell141.
ACCESSION AJ347134
VERSION AJ347134.1 GI:19773037
KEYWORDS IGWH gene; immunoglobulin heavy chain; variable region.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Brauning, A., Spieker, T., Willenbrock, K., Gaulard, P., Wacker, H.H., Rajewsky, K., Hansmann, M.L. and Kuppers, R.
TITLE Survival and clonal expansion of mutating 'forbidden' (immunoglobulin receptor-deficient) epstein-barr virus-infected b cells in angioimmunoblastic t cell lymphoma
JOURNAL J. Exp. Med. 194 (7), 927-940 (2001)
MEDLINE 21464858
PUBMED 11581315
REFERENCE 2 (bases 1 to 229)
AUTHORS Brauning, A.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2001) Brauning, A., Pathology, University of Frankfurt, Theodor-Stern-Kai 7, 60590, GERMANY
FEATURES
source
1..229
location/Qualifiers
/organism="Homo sapiens"
/isolate="case2-cell141"
/db_xref="taxon:9606"
/rearranged
1..229
gene
/gene="IGWH4-61"
<1..>229
CDS
/gene="IGWH4-61"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="CAC87534.1"
/db_xref="GI:19773038"
/translation="GSYVSWIRQPAKGLWIGRLYTSNTNPNPSLKSRYTMSVDTSKQFSLKLSVTAADTAVYCCARDIGLDSW"
<1..>229
V_region
/gene="IGWH4-61"
/product="immunoglobulin heavy chain variable region"
BASE COUNT 51 a 66 c 65 g 47 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 229;
Best Local Similarity 100.0%; Pred. No. 6,1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGAGC 15
|||||
Db 4 AGTTACTACTGAGC 18
RESULT 9
HSVH4GL11
LOCUS HSVH4GL11 230 bp DNA linear PRI 27-JUN-1996
DEFINITION H.sapiens germline immunoglobulin heavy chain VH gene (clone VH4-GL11).
ACCESSION Z75348
VERSION Z75348.1 GI:1418949
KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; immunoglobulin variable region.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 230)
AUTHORS Voswinkel, J., Truemper, L., Carbon, G., Hopf, T., Pfreundschuh, M. and Gause, A.
TITLE Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with rheumatoid

arthritis
JOURNAL unpublished
REFERENCE 2 (bases 1 to 230)
AUTHORS Voswinkel, J.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland Medical School, D-66421 Homburg/Saar, GERMANY
FEATURES
source
1..230
location/Qualifiers
/organism="Homo sapiens"
/isolate="patient 8 with rheumatoid arthritis"
/db_xref="taxon:9606"
/chromosome="14"
/clone="VH4-GL11"
/cell_type="B lymphocyte"
/tissue_type="synovial membrane"
/clone_lib="VH4-GL"
/germline
V_segment 1..230
BASE COUNT 54 a 64 c 64 g 48 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 230;
Best Local Similarity 100.0%; Pred. No. 6,1e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGTTACTACTGAGC 15
|||||
Db 25 AGTTACTACTGAGC 39
RESULT 10
HSVH4GL10
LOCUS HSVH4GL10 234 bp DNA linear PRI 27-JUN-1996
DEFINITION H.sapiens germline immunoglobulin heavy chain VH gene (clone VH4-GL10).
ACCESSION Z75347
VERSION Z75347.1 GI:1418948
KEYWORDS immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; immunoglobulin variable region.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 234)
AUTHORS Voswinkel, J., Truemper, L., Carbon, G., Hopf, T., Pfreundschuh, M. and Gause, A.
TITLE Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with rheumatoid arthritis
JOURNAL unpublished
REFERENCE 2 (bases 1 to 234)
AUTHORS Voswinkel, J.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland Medical School, D-66421 Homburg/Saar, GERMANY
FEATURES
source
1..234
location/Qualifiers
/organism="Homo sapiens"
/isolate="patient 8 with rheumatoid arthritis"
/db_xref="taxon:9606"
/chromosome="14"
/clone="VH4-GL10"
/cell_type="B lymphocyte"
/tissue_type="synovial membrane"
/clone_lib="VH4-GL"
/germline
V_segment 1..234
BASE COUNT 56 a 65 c 65 g 48 t
ORIGIN
Query Match 100.0%; Score 15; DB 9; Length 234;
Best Local Similarity 100.0%; Pred. No. 6,1e+02;

Matches	15;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	AGTTACTACTGGAGC	15						
Db	25	AGTTACTACTGGAGC	39						
RESULT 11									
LOCUS	HSVH4GL9		234 bp	DNA	linear	PRI 27-JUN-1996			
DEFINITION	H.sapiens germline immunoglobulin heavy chain VH gene (clone VH4-GL9).								
ACCESSION	Z75361								
VERSION	Z75361.1	GI:1418962							
KEYWORDS	immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; immunoglobulin variable region.								
SOURCE	Homo sapiens.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
AUTHORS	1 (bases 1 to 234)								
	Voswinkel,J., Truemper,L., Carbon,G., Hopf,T., Pfreundschuh,M. and Gause,A.								
TITLE	Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with rheumatoid arthritis								
JOURNAL	Unpublished								
REFERENCE	2 (bases 1 to 234)								
AUTHORS	Voswinkel,J.								
TITLE	Direct Submission								
JOURNAL	Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland Medical School, D-66421 Homburg/Saar, GERMANY								
FEATURES	location/Qualifiers								
source	1..234								
	/organism="Homo sapiens"								
	/isolate="patient 8 with rheumatoid arthritis"								
	/db_xref="taxon:9606"								
	/chromosome="14"								
	/clone="VH4-GL9"								
	/cell_type="B lymphocyte"								
	/tissue_type="synovial membrane"								
	/clone_lib="VH4-GL"								
	/germline								
	1..234								
V_segment	56	a	65	c	65	g	48	t	
BASE COUNT									
ORIGIN									
Query Match			100.0%;	Score 15;	DB 9;	length 234;			
Best Local Similarity			100.0%;	Pred. No. 6.1e+02;					
Matches	15;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	AGTTACTACTGGAGC	15						
Db	25	AGTTACTACTGGAGC	39						
RESULT 12									
LOCUS	HSVH4GL1		235 bp	DNA	linear	PRI 27-JUN-1996			
DEFINITION	H.sapiens immunoglobulin heavy chain VH gene (clone VH4-GL1).								
ACCESSION	Z75346								
VERSION	Z75346.1	GI:1418947							
KEYWORDS	immunoglobulin; immunoglobulin heavy chain; immunoglobulin superfamily; immunoglobulin variable region.								
SOURCE	Homo sapiens.								
ORGANISM	Homo sapiens								
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.								
AUTHORS	1 (bases 1 to 235)								
	Voswinkel,J., Truemper,L., Carbon,G., Hopf,T., Pfreundschuh,M. and Gause,A.								
TITLE	Evidence for a selected humoral immune response encoded by VH4 family genes in the synovial membrane of a patient with rheumatoid								

JOURNAL	Unpublished	arthritis
REFERENCE	2 (bases 1 to 235)	
AUTHORS	Voswinkel, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (14-JUN-1996) Voswinkel J., Internal Medicine I, Saarland Medical School, D-66421 Homburg/Saar, GERMANY	
FEATURES	Location/Qualifiers	
Source	1..235	
	/organism="Homo sapiens"	
	/isolate="patient 8 with rheumatoid arthritis"	
	/db_xref="taxon:9606"	
	/chromosome="14"	
	/clone="VH4-GL1"	
	/cell_type="B lymphocyte"	
	/tissue_type="synovial membrane"	
	/clone_1lb="VH4-GL"	
	/germline	
V-segment	1..235	
BASE COUNT	55 a 64 c 67 g 49 t	
ORIGIN		
Query Match	100.0%;	Score 15; DB 9; Length 235;
Best Local Similarity	100.0%;	Pred. No. 6e+02;
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AGTTACTACTGGAGC 15	
Db	26 AGTTACTACTGGAGC 40	
RESULT 13		
HSIGH2B1	236 bp	DNA linear PRI 07-JUL-1998
LOCUS	H.sapiens immunoglobulin rearranged heavy chain gene DP66 (clone 2B1).	
DEFINITION	271777.1 GI:1279486	
ACCESSION	271777	
VERSION	271777.1	
KEYWORDS	diversity region; immunoglobulin; immunoglobulin heavy chain; joining region; variable region.	
SOURCE	Homo sapiens.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
AUTHORS	1 (bases 1 to 236)	
TITLE	Dunn-Walters, D.K.	
JOURNAL	Direct Submission	
	Submitted (19-APR-1996) Dunn-Walters D.K., University College Medical School, Department of Histopathology, Rockefeller Building, University Street, London, WC1E 6JJ, UK	
REFERENCE	2 (bases 1 to 236)	
AUTHORS	Dunn-Walters, D.K., Boursier, L., Spencer, J. and Isaacson, P.G.	
TITLE	Analysis of immunoglobulin genes in splenic marginal zone lymphoma suggests ongoing mutation	
JOURNAL	Hum. Pathol. 29 (6), 585-593 (1998)	
MEDLINE	98297601	
PUBMED	9635678	
FEATURES	Location/Qualifiers	
Source	1..236	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="14"	
	/clone="2B1"	
	/tissue_type="spleen (splenic marginal zone lymphoma)"	
	/rearranged	
BASE COUNT	51 a 69 c 63 g 53 t	
ORIGIN		
Query Match	100.0%;	Score 15; DB 9; Length 236;
Best Local Similarity	100.0%;	Pred. No. 6e+02;
Matches	15; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 AGTTACTACTGGAGC 15	

Db 43 AGTTACTACTGAGC 57

RESULT 14
HSBDP71RB 237 bp DNA linear PRI 03-OCT-1996
LOCUS H.sapiens B9-g1DP71RB germline gene for immunoglobulin heavy chain
DEFINITION variable region.
X87091
X87091.1 GI:1592730
germline gene; immunoglobulin; immunoglobulin heavy chain; variable region.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 237)
AUTHORS Brezinschek,H.P., Brezinschek,R.I. and Lipsky,P.E.
TITLE Analysis of the heavy chain repertoire of human peripheral B cells using single-cell polymerase chain reaction
JOURNAL J. Immunol. 155 (1), 190-202 (1995)
MEDLINE 95325588
PUBMED 7602095
REFERENCE 2 (bases 1 to 237)
AUTHORS Brezinschek,H.P.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1995) Brezinschek H.P., Department of Internal Medicine, Harold C. Simmons Arthritis Research Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-8884, USA
COMMENT On Oct 4, 1996 this sequence version replaced gi:1052697.
FEATURES
source location/Qualifiers
1..237
/organism="Homo sapiens"
/isolate="donor RB"
/db_xref="taxon:9606"
/chromosome="14"
/cell_type="granulocyte"
/tissue_type="blood"
1..237
/gene="B9-g1DP71RB"
1..237
/gene="B9-g1DP71RB"
/note="VH4 family germline gene, polymorphic form of DP-71"
BASE COUNT 54 a 70 c 61 g 50 t 2 others
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 237;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
Db 37 AGTTACTACTGAGC 51

RESULT 15
HSA347131 238 bp DNA linear PRI 25-MAR-2002
LOCUS HSA347131
DEFINITION Homo sapiens partial IGVH4-61 gene for immunoglobulin heavy chain
variable region, isolate case2-cell1150.
AJ347131
AJ347131.1 GI:19773031
IGVH gene; immunoglobulin heavy chain; variable region.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Brauning,A., Spieker,T., Willenbrock,K., Gaulard,P., Wacker,H.H., Rajewsky,K., Hansmann,M.L. and Kuppers,R.

TITLE Survival and clonal expansion of mutating 'forbidden' (immunoglobulin receptor-deficient) epstein-barr virus-infected b cells in angioimmunoblastic t cell lymphoma
JOURNAL J. Exp. Med. 194 (7), 927-940 (2001)
MEDLINE 21464858
PUBMED 11581315
REFERENCE 2 (bases 1 to 238)
AUTHORS Braeuninger,A.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2001) Braeuninger A., Pathology, University of Frankfurt, Theodor-Stern-Kai 7, 60590, GERMANY
FEATURES
source location/Qualifiers
1..238
/organism="Homo sapiens"
/isolate="case2-cell1150"
/db_xref="taxon:9606"
/rearranged
1..238
/gene="IGVH4-61"
<1..>238
/gene="IGVH4-61"
/codon_start=1
/product="immunoglobulin heavy chain variable region"
/protein_id="CAC87531.1"
/db_xref="GI:19773032"
/translation="SYWSWIRQPAKGLEWIGRIYTSGSTNYKSLKSRVTMSVDT
KNQFSKLSSVTAADTAVYYCARGTGLRWAFDIW"
<1..>238
/gene="IGVH4-61"
/product="immunoglobulin heavy chain variable region"
BASE COUNT 54 a 61 c 74 g 49 t
ORIGIN

Query Match 100.0%; Score 15; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
Db 1 AGTTACTACTGAGC 15

Search completed: June 26, 2003, 03:47:37
Job time : 425.6 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 26, 2003, 03:00:35 ; Search time 86 Seconds
(without alignments)
392.790 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agttactactgagac 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	15	100.0	15	20	AAx32408	Ab6 variable heavy
2	15	100.0	110	24	ABK71391	Thrombopoietin (TP
3	15	100.0	171	22	ABA70643	Human foetal liver
4	15	100.0	171	22	ABA37202	Probe #15668 for g
5	15	100.0	171	22	AAK18889	Human brain expres
6	15	100.0	171	22	AAK44833	Human bone marrow
7	15	100.0	171	22	AAI25038	Probe #14971 for g
8	15	100.0	171	22	AAI50810	Probe #19496 used
9	15	100.0	171	24	ABS19072	Human genome-deriv

10	15	100.0	281	22	ABA70163	Human foetal liver
11	15	100.0	281	22	ABA36928	Probe #15394 for g
12	15	100.0	281	22	AAK18383	Human brain expres
13	15	100.0	281	22	AAK44283	Human bone marrow
14	15	100.0	281	22	AAI24810	Probe #14743 for g
15	15	100.0	281	22	AAI50286	Probe #18972 used
16	15	100.0	281	24	ABS18519	Human genome-deriv
17	15	100.0	319	22	AAK20407	Human brain expres
18	15	100.0	342	24	ABK71388	Thrombopoietin (TP
19	15	100.0	348	14	AAQ42697	Vh 71-4. Homo sap
20	15	100.0	348	14	AAQ42699	VH411. Homo sapie
21	15	100.0	348	14	AAQ42700	VH415. Homo sapie
22	15	100.0	348	14	AAQ42701	VH416. Homo sapie
23	15	100.0	357	18	AAT72126	CEA-specific antib
24	15	100.0	363	22	AAH42398	Nucleotide sequenc
25	15	100.0	363	24	ABA94218	ebvH1qM MSI19D10 h
26	15	100.0	369	22	AAI68756	Human autoantibody
27	15	100.0	378	22	AAH42392	Nucleotide sequenc
28	15	100.0	397	20	AAZ24417	Human bladder tumo
29	15	100.0	408	18	AAT79919	Immunoglobulin rB6
30	15	100.0	417	21	AAAI3938	Human PTHrP monocl
31	15	100.0	426	24	ABK71396	DNA encoding throm
32	15	100.0	456	21	AAZ42290	Human 5' EST isola
33	15	100.0	462	22	ABA57539	Human foetal liver
34	15	100.0	462	22	ABA27008	Probe #5474 for ge
35	15	100.0	462	22	AAK05586	Human brain expres
36	15	100.0	462	22	AAK31189	Human bone marrow
37	15	100.0	462	22	AAI15585	Probe #5518 for ge
38	15	100.0	462	22	AAI37094	Probe #5780 used t
39	15	100.0	462	24	ABS05941	Human genome-deriv
40	15	100.0	470	22	ABA58037	Human foetal liver
41	15	100.0	470	22	ABA27294	Probe #5760 for ge
42	15	100.0	470	22	AAK06111	Human brain expres
43	15	100.0	470	22	AAK31762	Human bone marrow
44	15	100.0	470	22	AAI15819	Probe #5752 for ge
45	15	100.0	470	22	AAI37638	Probe #6324 used t

ALIGNMENTS

RESULT 1			
AAx32408			
ID	AAx32408 standard; DNA; 15 BP.		
XX			
AC	AAx32408;		
XX			
DT	17-JUN-1999 (first entry)		
XX			
DE	Ab6 variable heavy (VH) chain CDR1 encoding DNA.		
XX			
KW	Agonist antibody; thrombopoietin receptor; TP0-R; thrombopoietin; DIC;		
KW	megakaryocyte; platelet; immunological; hematopoietic; thrombocytopenia;		
KW	bone marrow hypoplasia; disseminated intravascular coagulation; anemia;		
KW	myelodysplasia; myelotoxic chemotherapy; leukaemia; tumour; MusK; CDR;		
KW	neuromuscular; muscular dystrophy; complementarity determining region;		
XX	variable heavy chain; variable light chain; VH; VL; SS.		
OS			
XX	Homo sapiens.		
PN	WO9910494-A2.		
XX			
PD	04-MAR-1999.		
XX			
PF	21-AUG-1998; 98WO-US17364.		
XX			
PR	25-AUG-1997; 97US-0918148.		
XX			
PA	(GETH) GENENTECH INC.		
XX			
PI	Adams CW, Carter PJ, Fendly BM, Gurney AL;		
XX			
DR	WPI; 1999-204666/17.		

DR P-PSDB; AAY06707.
XX
PT New thrombopoietin receptor agonist antibodies - useful for
PT treating immunological or hematological disorders
XX
PS Claim 10; Page 81; 86pp; English.
XX
CC The invention relates to an agonist antibody (Ab) which binds to a
CC thrombopoietin receptor (TPO-R). The antibodies which bind the TPO-R can
CC be used in the same way and for the same indications as thrombopoietin
CC (TPO). They can stimulate proliferation, differentiation or growth of
CC megakaryocytes. They may also be able to stimulate megakaryocytes to
CC increase platelet production. They can be used for treating
CC immunological or hematopoietic disorders, especially thrombocytopenia.
CC Thrombocytopenia -associated bone marrow hypoplasia (e.g. aplastic anemia
CC following chemotherapy or bone marrow transplant) may be effectively
CC treated with the antibody compounds as well as disorders such as
CC disseminated intravascular coagulation (DIC), immune thrombocytopenia
CC (HIV-induced and non HIV-induced), chronic idiopathic thrombocytopenia,
CC congenital thrombocytopenia, thrombotic thrombocytopenia and
CC myelodysplasia. They can also be used in e.g. myelotoxic chemotherapy for
CC treatment of solid tumours or leukaemia, myeloablative chemotherapy for
CC autologous or allogeneic bone marrow transplant, myelodysplasia,
CC idiopathic aplastic anemia, congenital thrombocytopenia, and immune
CC thrombocytopenia. The antibodies which bind to the MUSK receptor can be
CC used for improving neuromuscular function in a patient, e.g. in muscular
CC dystrophy. The products can also be used for detection and diagnosis. The
CC antibodies have a longer half-life than the natural ligand for the TPO-R.
CC Sequences AAX32387-X32413 represent DNA fragments encoding the CDR1,
CC CDR2, and CDR3 regions of variable heavy (VH) chains and variable light
CC (VL) chains of antibodies Ab1 to Ab6.
XX
SQ Sequence 15 BP; 4 A; 3 C; 4 G; 4 T; 0 other;

Query Match 100.0%; Score 15; DB 20; Length 15;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
Db 1 AGTTACTACTGAGC 15

RESULT 2
ABK71391/C
ID ABK71391 standard; DNA; 110 BP.
XX
AC ABK71391;
XX
DT 30-JUL-2002 (first entry)
XX
DE Thrombopoietin (TPO) agonist antibody associated polynucleotide #12.
XX
KW Modified antibody; thrombopoietin; TPO; agonist;
KW TPO receptor; platelet reduction-associated blood disease;
KW thrombocytopenia; cancer chemotherapy; leukaemia; signal transduction;
KW ds.
XX
OS Synthetic.
XX
PN WO200233072-A1.
XX
PD 25-APR-2002.
XX
PF 22-OCT-2001; 2001WO-JP09259.
XX
PR 20-OCT-2000; 2000JP-0321821.
PR 17-APR-2001; 2001WO-JP03288.
PR 12-SEP-2001; 2001JP-0277314.
XX
PA (CHUS) CHUGAI SEIYAKU KK.
XX
PI Tsuchiya M, Ohtomo T, Yabuta N, Tsunoda H, Orita T;

XX
DR WPI; 2002-383513/41.
XX
PT Degraded thrombopoietin agonist antibodies containing H and L chain V
PT domains of monoclonal antibody, useful in preventives and/or remedies
PT for blood diseases, thrombocytopenia following cancer chemotherapy or
PT leukaemia
XX
PS Example 8; Page 195; 213pp; Japanese.
XX
CC The invention describes a modified antibody comprising at least 2 heavy
CC chain variable domains and 2 or more light chain variable domains of an
CC antibody, and exhibits thrombopoietin (TPO) agonistic effect by causing
CC the TPO receptor to crosslink. The antibodies are useful in preventives
CC and/or remedies for platelet reduction-associated blood diseases,
CC thrombocytopenia following cancer chemotherapy or leukaemia. The
CC antibody can act as a TPO signal transduction agonist by transducing a
CC signal into cells by crosslinking a TPO receptor to exert TPO agonism.
CC This sequence represents thrombopoietin (TPO) agonist antibody associated
CC polynucleotide used in the creation of the modified antibody described in
CC the invention.
XX
SQ Sequence 110 BP; 26 A; 34 C; 31 G; 19 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 110;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15;
Db 58 AGTTACTACTGAGC 44

RESULT 3
ABA70643/C
ID ABA70643 standard; DNA; 171 BP.
XX
AC ABA70643;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #18948.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 4; SEQ ID NO 18948; 639pp + sequence listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
 |||||
Db 82 AGTTACTACTGGAGC 68

RESULT 4
ABA37202/c
ID ABA37202 standard; DNA; 171 BP.
XX
AC ABA37202;
XX
DT 23-JAN-2002 (first entry)
XX
DE Probe #15668 for gene expression analysis in human heart cell sample.
XX
KW Human; gene expression; heart; microarray; vascular system; probe;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease; ss.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 4; SEQ ID No 15668; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart. The
CC present sequence is one such probe. The probes may be used for
CC predicting, measuring and displaying gene expression in samples derived
CC from the human heart via microarrays. By measuring gene expression, the
CC probes are useful for predicting, diagnosing, grading, staging,
CC monitoring and prognosing diseases of the human heart and vascular system
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and
CC congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
 |||||
Db 82 AGTTACTACTGGAGC 68

RESULT 5
AAK18889/c
ID AAK18889 standard; DNA; 171 BP.
XX
AC AAK18889;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe SEQ ID NO: 18880.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 18880; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
 |||||
Db 82 AGTTACTACTGGAGC 68

RESULT 6


```

AAK44833/c
ID AAK44833 standard; DNA; 171 BP.
XX
XX AAK44833;
AC
DT 06-NOV-2001 (first entry)
XX
XX Human bone marrow expressed single exon probe SEQ ID NO: 19390.
DE
XX Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
XX WO200157276-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00668.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX Penn SG, Hanzel DK, Chen W, Rank DR;
PI
XX WPI; 2001-488900/53.
DR
XX
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
PT
XX
XX Example 4; SEQ ID NO: 19390; 658bp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
CC
XX
XX Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;
SQ

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
   |||||||||||||||
Db 82 AGTTACTACTGGAGC 68

RESULT 7
AAI25038/c
ID AAI25038 standard; DNA; 171 BP.
XX
XX AAI25038;
AC
DT 12-OCT-2001 (first entry)
XX
XX Probe #14971 for gene expression analysis in human cervical cell sample.
DE
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157278-A2.
PN

```

[illegible]

PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
PA
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID No 19496; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
Db 82 AGTTACTACTGGAGC 68
|||||

RESULT 9
ABS19072/c
ID ABS19072 standard; DNA; 171 BP.
XX
AC ABS19072;
XX
DT 19-AUG-2002 (first entry)
XX
DE Human genome-derived single exon probe ORF from lung SEQ ID No 19063.
XX
KW Human; 'ds; single exon probe; asthma; lung cancer; COPD; ILD;
KW chronic obstructive pulmonary disease; interstitial lung disease;
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagenen syndrome;
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
KW primary ciliary dyskinesia; pulmonary hypertension;
KW hyaline membrane disease; open reading frame; ORF.
XX
OS Homo sapiens.
XX
PN WO200186003-A2.
XX
PD 15-NOV-2001.
XX
PF 30-JAN-2001; 2001WO-US00665.
XX
PR 04-FEB-2000; 2000US-180312P.
PR 26-MAY-2000; 2000US-207456P.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-234687P.
PR 27-SEP-2000; 2000US-236359P.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2002-114183/15.
XX
PT Spatially-addressable set of single exon nucleic acid probes, used to

PT measure gene expression in human lung samples -
XX
XX Claim 4; SEQ ID No 19063; 634pp; English.
PS
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human lung comprising single exon nucleic acid probes having one of
CC 12614 nucleic acid sequences mentioned in the specification, or their
CC complements or the 12387 open reading frames derived from the 12614
CC probes. Also included are a microarray comprising the novel set of
CC probes; the novel set of probes which hybridise at high stringency to a
CC nucleic acid expressed in the human lung; measuring gene expression in a
CC sample derived from human lung, comprising (a) contacting the array with
CC a collection of detectably labeled nucleic acids derived from human lung
CC mRNA, and (b) measuring the label detectably bound to each probe of
CC the array; identifying exons in a eukaryotic genome, comprising
CC (a) algorithmically predicting at least one exon from genomic sequences
CC of the eukaryote; and (b) detecting specific hybridisation of detectably
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,
CC having a fragment identical to the predicted exon, the probe is included
CC in the above mentioned microarray; assigning exons to a single gene,
CC comprising (a) identifying exons from genomic sequence by the method
CC above and (b) measuring the expression of each of the exons in several
CC tissues and/or cell types using hybridisation to a single exon
CC microarrays having a probe with the exon, where a common pattern of
CC expression of the exons in the tissues and/or cell types indicates that
CC the exons should be assigned to a single gene; a peptide comprising one
CC of 12011 sequences, mentioned in the specification, or encoded by the
CC probes/open reading frames (ORF). The probes are used for gene
CC expression analysis, and for identifying exons in a gene, particularly
CC using human lung derived mRNA and for the study of lung diseases
CC such as asthma, lung cancer, chronic obstructive pulmonary disease
CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
CC pulmonary alveolar proteinosis, Karagenen syndrome, fibrocystic
CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
CC and hyaline membrane disease. The present sequence is a single exon
CC probe open reading frame of the invention.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct-sequences.
XX
SQ Sequence 171 BP; 51 A; 47 C; 37 G; 36 T; 0 other;

Query Match 100.0%; Score 15; DB 24; Length 171;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
Db 82 AGTTACTACTGGAGC 68
|||||

RESULT 10
ABA70163
ID ABA70163 standard; DNA; 281 BP.
XX
AC ABA70163;
XX
DT 01-FEB-2002 (first entry)
XX
DE Human foetal liver single exon nucleic acid probe #18468.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.

XX	30-JAN-2001; 2001WO-US00669.
Pf	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-483447/52.
PT	Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human fetal liver -
PS	
XX	Claim 4; SEQ ID NO 18468; 639pp + sequence listing; English.
CC	The invention relates to a single exon nucleic acid probe for measuring human gene expression in a sample derived from human foetal liver. The single exon nucleic acid probes may be used for predicting, measuring and displaying gene expression in samples derived from human fetal liver. The present sequence is a single exon nucleic acid probe of the invention.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
SQ	Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;
QY	Query Match 100.0%; Score 15; DB 22; Length 281; Best Local Similarity 100.0%; Pred. No. 84; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db	1 AGTTACTACTGGAGC 15 64 AGTTACTACTGGAGC 78
ID	RESULT 11 ABA36928
XX	ABA36928 standard; DNA;.281 BP.
AC	ABA36928;
DT	23-JAN-2002 (first entry)
DE	Probe #15394 for gene expression analysis in human heart cell sample.
XX	
KX	Human; gene expression; heart; microarray; vascular system; probe;
KW	cardiovascular disease; hypertension; cardiac arrhythmia;
KW	congenital heart disease; ss.
OS	Homo sapiens.
XX	
PN	WO200157274-A2.
XX	
PD	09-AUG-2001.
XX	
PF	30-JAN-2001; 2001WO-US00666.
XX	
PR	04-FEB-2000; 2000US-0180312.
PR	26-MAY-2000; 2000US-0207456.
PR	30-JUN-2000; 2000US-0608408.
PR	03-AUG-2000; 2000US-0632366.
PR	21-SEP-2000; 2000US-0234687.
PR	27-SEP-2000; 2000US-0236359.
PR	04-OCT-2000; 2000GB-0024263.
XX	

PA	(MOLE-) MOLECULAR DYNAMICS INC.
XX	
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
XX	
DR	WPI; 2001-488899/53.
XX	
PT	Single exon nucleic acid probes for analyzing gene expression in human hearts -
PS	Claim 4; SEQ ID No 15394; 530pp; English.
XX	
CC	The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart. The present sequence is one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease.
CC	Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
SO	Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;
QY	1 AGTTACTACTGGAGC 15 1111111111
Db	64 AGTTACTACTGGAGC 78
RESULT 12	
AAK18383	
ID	AAK18383 standard; DNA; 281 BP.
AC	AAK18383;
DT	05-NOV-2001 (first entry)
DE	Human brain expressed single exon probe SEQ ID NO: 18374.
XX	
KW	Human; brain expressed exon; gene expression analysis; probe; microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer; ss.
OS	Homo sapiens.
PN	WO200157275-A2.
PD	09-AUG-2001.
PF	30-JAN-2001; 2001WO-US00667.
PR	04-FEB-2000; 2000US-0180312. 26-MAY-2000; 2000US-0207456. 30-JUN-2000; 2000US-0608408. 03-AUG-2000; 2000US-0632365. 21-SEP-2000; 2000US-0234687. 27-SEP-2000; 2000US-0236359. 04-OCT-2000; 2000GB-0024263.
PA	(MOLE-) MOLECULAR DYNAMICS INC.
PI	Penn SG, Hanzel DK, Chen W, Rank DR;
DR	WPI; 2001-483446/52.
PT	Single exon nucleic acid probes for analyzing gene expression in human brains -

```
XX Example 4; SEQ ID NO: 18374; 650pp + Sequence Listing; English.
PS
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is one of the probes of the
CC invention.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match      100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||||||
Db      64 AGTTACTACTGAGC 78

RESULT 13
AAK44283
ID AAK44283 standard; DNA; 281 BP.
XX
AC AAK44283;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human bone marrow expressed single exon probe SEQ ID NO: 18840.
XX
KW Human; bone marrow expressed exon; gene expression analysis; probe;
KW microarray; cancer; leukaemia; lymphoma; myeloma; ss.
XX
OS Homo sapiens.
XX
PN WO200157276-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00668.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488900/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human bone marrow -
XX
XX
PS Example 4; SEQ ID NO: 18840; 658pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC bone marrow. They can be used to measure gene expression in bone marrow
CC samples, which may enable the improved diagnosis and treatment of cancers
CC such as lymphoma, leukaemia and myeloma. The present sequence is one of
CC the probes of the invention.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match      100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
```

```
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||||||
Db      64 AGTTACTACTGAGC 78

RESULT 14
AAI24810
ID AAI24810 standard; DNA; 281 BP.
XX
AC AAI24810;
XX
DT 12-OCT-2001 (first entry)
XX
DE Probe #14743 for gene expression analysis in human cervical cell sample.
XX
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer; ss.
XX
OS Homo sapiens.
XX
PN WO200157278-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00670.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488901/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -
XX
PS Claim 25; SEQ ID NO 14743; 487pp; English.
XX
CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match      100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||||||
Db      64 AGTTACTACTGAGC 78

RESULT 15
AAI50286
ID AAI50286 standard; DNA; 281 BP.
XX
```

AC AAI50286;
XX
DT 17-OCT-2001 (first entry)
XX
DE Probe #18972 used to measure gene expression in human placenta sample.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder; ss.
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00663.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488897/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX
PS Claim 25; SEQ ID NO 18972; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.
XX
SQ Sequence 281 BP; 61 A; 81 C; 84 G; 55 T; 0 other;

Query Match 100.0%; Score 15; DB 22; Length 281;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
Db 64 AGTTACTACTGGAGC 78

Search completed: June 26, 2003, 03:11:51
Job time : 88 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:03:30 ; Search time 19.2 Seconds
(without alignments)
239.591 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agttactactggagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	100.0	15	4	US-08-918-148-42
2	15	100.0	357	2	US-08-652-816A-20
3	15	100.0	622	2	US-08-545-809A-59
4	15	100.0	650	3	US-08-545-809A-4
5	15	100.0	1212	3	US-08-545-809A-61
6	14	93.3	15	2	US-08-477-553A-3
7	14	93.3	15	2	US-08-477-553A-6
8	14	93.3	43	4	US-09-042-353-397
9	14	93.3	43	4	US-09-042-353-400
10	14	93.3	43	4	US-08-758-417A-250
11	14	93.3	43	4	US-08-758-417A-247
12	14	93.3	243	4	US-09-042-353-148
13	14	93.3	243	4	US-08-758-417A-412
14	14	93.3	282	4	US-09-042-353-149
15	14	93.3	282	4	US-08-758-417A-413
16	14	93.3	285	4	US-09-042-353-150
17	14	93.3	285	4	US-08-758-417A-414
18	14	93.3	297	4	US-09-042-353-151
19	14	93.3	297	4	US-08-758-417A-415
20	14	93.3	321	2	US-08-477-553A-47
21	14	93.3	363	2	US-08-477-553A-50
22	14	93.3	369	3	US-08-793-450-3
23	14	93.3	399	3	US-08-724-752-10
24	14	93.3	403	4	US-09-042-353-357
25	14	93.3	403	4	US-08-758-417A-205
26	14	93.3	404	4	US-09-042-353-355
27	14	93.3	404	4	US-08-758-417A-203

28	14	93.3	524	4	US-09-042-353-419	Sequence 419, App
29	14	93.3	524	4	US-08-758-417A-219	Sequence 219, App
30	14	93.3	631	3	US-08-545-809A-31	Sequence 31, Appl
31	14	93.3	687	3	US-08-545-809A-34	Sequence 34, Appl
32	14	93.3	1418	4	US-08-793-450-7	Sequence 7, Appl
33	14	93.3	1567	3	US-09-049-672A-17	Sequence 17, Appl
34	14	93.3	2624	4	US-09-370-838-19	Sequence 19, Appl
35	14	93.3	4926	4	US-09-042-353-418	Sequence 418, App
36	14	93.3	4926	4	US-08-758-417A-268	Sequence 268, App
37	13.4	89.3	15	2	US-08-477-553A-2	Sequence 2, Appl
38	13.4	89.3	30	1	US-08-360-125-37	Sequence 37, Appl
39	13.4	89.3	30	2	US-08-450-578-37	Sequence 37, Appl
40	13.4	89.3	30	2	US-09-017-628-37	Sequence 37, Appl
41	13.4	89.3	30	2	US-09-014-880-37	Sequence 37, Appl
42	13.4	89.3	30	4	US-08-450-363-37	Sequence 37, Appl
43	13.4	89.3	354	2	US-08-652-816A-23	Sequence 23, Appl
44	13.4	89.3	366	1	US-08-360-125-9	Sequence 9, Appl
45	13.4	89.3	366	2	US-08-450-578-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-918-148-42
; Sequence 42, Application US/08918148A
; Patent No. 6342220
; GENERAL INFORMATION:
; APPLICANT: Adams, Camellia
; APPLICANT: W.
; APPLICANT: Carter, Paul J.
; APPLICANT: Fendly, Brian M.
; APPLICANT: Gurney, Austin L.
; TITLE OF INVENTION: Agonist Antibodies
; FILE REFERENCE: P0979
; CURRENT APPLICATION NUMBER: US/08/918,148A
; CURRENT FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 79
; SEQ ID NO 42
; LENGTH: 15
; TYPE: DNA
; ORGANISM: artificial
; FEATURE:
; NAME/KEY: 12E10scFv VH CDRI
; LOCATION: 1-15
; OTHER INFORMATION:
; US-08-918-148-42

Query Match 100.0%; Score 15; DB 4; Length 15;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
Db 1 AGTTACTACTGGAGC 15

RESULT 2
US-08-652-816A-20
; Sequence 20, Application US/08652816A
; Patent No. 5872215
; GENERAL INFORMATION:
; APPLICANT: Osbourn, JK
; APPLICANT: Allen, DJ
; APPLICANT: McCafferty, JG
; TITLE OF INVENTION: Specific binding members, materials and
; TITLE OF INVENTION: methods.
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Tooley, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois

```

; COUNTRY: United States of America
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,816A
; FILING DATE: 23-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9125579.4
; FILING DATE: 02-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9125579.8
; FILING DATE: 02-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9206318.9
; FILING DATE: 24-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9206372.6
; FILING DATE: 23-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525004.9
; FILING DATE: 07-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9610824.6
; FILING DATE: 23-MAY-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB92/02240
; FILING DATE: 02-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/244,597
; FILING DATE: 01-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: David W. Clough
; REGISTRATION NUMBER: 36,107
; REFERENCE/DOCKET NUMBER: 28111/33308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 357 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-652-816A-20

Query Match          100.0%; Score 15; DB 2; Length 357;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
   |||
Db 91 AGTTACTACTGGAGC 105

RESULT 3
US-08-545-809A-59
; Sequence 59, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 59:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 622 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: human lymphoblast
; CELL LINE: CGM1
;
US-08-545-809A-59

Query Match          100.0%; Score 15; DB 3; Length 622;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGGAGC 15
   |||
Db 381 AGTTACTACTGGAGC 395

RESULT 4
US-08-545-809A-4
; Sequence 4, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001

```

```
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 650 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: human lymphoblast
; CELL LINE: CGM1
; US-08-545-809A-4

Query Match      100.0%; Score 15; DB 3; Length 650;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
Db      319 AGTTACTACTGAGC 333

RESULT 5
US-08-545-809A-61
; Sequence 61, Application US/08545809A
; Patent No. 6096878
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Matsuda, Fumihiko
; TITLE OF INVENTION: HUMAN IMMUNOGLOBULIN VH GENE
; TITLE OF INVENTION: SEGMENTS AND DNA FRAGMENTS CONTAINING THE SAME
; NUMBER OF SEQUENCES: 145
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,809A
; FILING DATE: 27-MAR-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/00603
; FILING DATE: 10-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 06501/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 61:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1212 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CELL TYPE: human lymphoblast
; CELL LINE: CGM1
```

```
US-08-545-809A-61

Query Match      100.0%; Score 15; DB 3; Length 1212;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
Db      386 AGTTACTACTGAGC 400

RESULT 6
US-08-477-553A-3
; Sequence 3, Application US/08477553A
; Patent No. 5919910
; GENERAL INFORMATION:
; APPLICANT: HUGHES-JONES, Nevin C
; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,553A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/856,034
; FILING DATE: 23-JUNE-1992
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8925590.5
; FILING DATE: 13-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 007330-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-477-553A-3

Query Match      93.3%; Score 14; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GTTACTACTGAGC 15
Db      2 GTTACTACTGAGC 15

RESULT 7
US-08-477-553A-6
; Sequence 6, Application US/08477553A
; Patent No. 5919910
; GENERAL INFORMATION:
; APPLICANT: HUGHES-JONES, Nevin C
```

```

; TITLE OF INVENTION: MONOCLONAL ANTIBODIES
; NUMBER OF SEQUENCES: 55
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,553A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/856,034
; FILING DATE: 23-JUNE-1992
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: GB 8925590.5
; FILING DATE: 13-NOV-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M.
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 007330-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 15 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-477-553A-6

Query Match 93.3%; Score 14; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGGAGC 15
DB 2 GTTACTACTGGAGC 15

RESULT 8
US-09-042-353-397
; Sequence 397, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; APPLICANT: Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; TITLE OF INVENTION: Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/09/042,353
; FILING DATE: 13-MAR-1998
; CLASSIFICATION: 800
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/810,279
; FILING DATE: 17-DEC-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/853,408
; FILING DATE: 18-MAR-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/904,068
; FILING DATE: 23-JUN-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US96/16433
; FILING DATE: 10-OCT-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,417
; FILING DATE: 02-DEC-1996
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/21803
; FILING DATE: 01-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 014643-009040US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 397:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 43 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-09-042-353-397

Query Match 93.3%; Score 14; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGGAGC 15

```

DB 5 GTTACTACTGAGC 18

RESULT 9

US-09-042-353-400/c

; Sequence 400, Application US/09042353

; Patent No. 6255458

; GENERAL INFORMATION:

; APPLICANT: Lonberg, Nils

; APPLICANT: Kay, Robert M.

; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for

; TITLE OF INVENTION: Producing Heterologous Antibodies

; NUMBER OF SEQUENCES: 421

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/042,353

; FILING DATE: 13-MAR-1998

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/810,279

; FILING DATE: 17-DEC-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/853,408

; FILING DATE: 18-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/904,068

; FILING DATE: 23-JUN-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/990,860

; FILING DATE: 16-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/053,131

; FILING DATE: 26-APR-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/096,762

; FILING DATE: 22-JUL-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/155,301

; FILING DATE: 18-NOV-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,739

; FILING DATE: 03-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/165,699

; FILING DATE: 10-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/209,741

; FILING DATE: 09-MAR-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/352,322

; FILING DATE: 07-DEC-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/544,404

; FILING DATE: 10-OCT-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/728,463

; FILING DATE: 10-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/US96/16433

; FILING DATE: 10-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/758,417

; FILING DATE: 02-DEC-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/US97/21803

; FILING DATE: 01-DEC-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Apple, Randolph T.

; REGISTRATION NUMBER: 36,429

; REFERENCE/DOCKET NUMBER: 014643-009040US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 576-0200

; TELEFAX: (415) 576-0300

; INFORMATION FOR SEQ ID NO: 400:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 43 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA

US-09-042-353-400

Query Match 93.3%; Score 14; DB 4; Length 43;

Best Local Similarity 100.0%; Pred. No. 19;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGAGC 15

Db 16 GTTACTACTGAGC 3

RESULT 10

US-08-758-417A-247

; Sequence 247, Application US/08758417A

; Patent No. 6300129

; GENERAL INFORMATION:

; APPLICANT: Lonberg, Nils

; APPLICANT: Kay, Robert M.

; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for

; TITLE OF INVENTION: Producing Heterologous Antibodies

; NUMBER OF SEQUENCES: 417

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Townsend and Townsend and Crew LLP

; STREET: Two Embarcadero Center, Eighth Floor

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/758,417A

; FILING DATE: 02-DEC-1996

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/728,463

; FILING DATE: 10-OCT-1996

; APPLICATION NUMBER: US 08/544,404

; FILING DATE: 10-OCT-1995

; APPLICATION NUMBER: US 08/352,322

; FILING DATE: 07-DEC-1994

; APPLICATION NUMBER: US 08/209,741

; FILING DATE: 09-MAR-1994

; APPLICATION NUMBER: US 08/165,699

; FILING DATE: 10-DEC-1993

; APPLICATION NUMBER: US 08/161,739

; FILING DATE: 03-DEC-1993

; APPLICATION NUMBER: US 08/155,301

; FILING DATE: 18-NOV-1993

; APPLICATION NUMBER: US 08/096,762

; FILING DATE: 22-JUL-1993

; APPLICATION NUMBER: US 08/053,131

FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 247:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 247:
US-08-758-417A-247

Query Match 93.3%; Score 14; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGAGC 15
|||||
Db 5 GTTACTACTGAGC 18

RESULT 11

US-08-758-417A-250/c
Sequence 250, Application US/08758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131

FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 250:
SEQUENCE CHARACTERISTICS:
LENGTH: 43 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
SEQUENCE DESCRIPTION: SEQ ID NO: 250:
US-08-758-417A-250

Query Match 93.3%; Score 14; DB 4; Length 43;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GTTACTACTGAGC 15
|||||
Db 16 GTTACTACTGAGC 3

RESULT 12

US-09-042-353-148
Sequence 148, Application US/09042353
Patent No. 6255458
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 421
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US96/16433
; FILING DATE: 10-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/758,417
; FILING DATE: 02-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US97/21803
; FILING DATE: 01-DEC-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Apple, Randolph T.
; REGISTRATION NUMBER: 36,429
; REFERENCE/DOCKET NUMBER: 014643-009040US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 148:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-042-353-148

```

```

Query Match          93.3%; Score 14; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GTTACTACTGAGC 15
        |||||||
Db      44 GTTACTACTGAGC 57

```

```

RESULT 13
US-08-758-417A-412
; Sequence 412, Application US/08758417A
; Patent No. 6300129
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
; Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 417
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/758,417A
; FILING DATE: 02-DEC-1996
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/728,463
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US 08/544,404
; FILING DATE: 10-OCT-1995
; APPLICATION NUMBER: US 08/352,322
; FILING DATE: 07-DEC-1994
; APPLICATION NUMBER: US 08/209,741
; FILING DATE: 09-MAR-1994
; APPLICATION NUMBER: US 08/165,699
; FILING DATE: 10-DEC-1993
; APPLICATION NUMBER: US 08/161,739
; FILING DATE: 03-DEC-1993
; APPLICATION NUMBER: US 08/155,301
; FILING DATE: 18-NOV-1993
; APPLICATION NUMBER: US 08/096,762
; FILING DATE: 22-JUL-1993
; APPLICATION NUMBER: US 08/053,131
; FILING DATE: 26-APR-1993
; APPLICATION NUMBER: US 07/990,860
; FILING DATE: 16-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Serafini, Andrew T.
; REGISTRATION NUMBER: 41,303
; REFERENCE/DOCKET NUMBER: 014643-009030US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 412:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 412:
; US-08-758-417A-412

```

```

Query Match          93.3%; Score 14; DB 4; Length 243;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 GTTACTACTGAGC 15
        |||||||
Db      44 GTTACTACTGAGC 57

```

```

RESULT 14
US-09-042-353-149
; Sequence 149, Application US/09042353
; Patent No. 6255458
; GENERAL INFORMATION:
; APPLICANT: Lonberg, Nils
; Kay, Robert M.
; TITLE OF INVENTION: Transgenic No. 6255458-Human Animals for
; Producing Heterologous Antibodies
; NUMBER OF SEQUENCES: 421
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

```

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/042,353
FILING DATE: 13-MAR-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,279
FILING DATE: 17-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/853,408
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/904,068
FILING DATE: 23-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US96/16433
FILING DATE: 10-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/758,417
FILING DATE: 02-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US97/21803
FILING DATE: 01-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Apple, Randolph T.
REGISTRATION NUMBER: 36,429
REFERENCE/DOCKET NUMBER: 014643-009040US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 149:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
US-09-042-353-149

```

Query Match 93.3%; Score 14; DB 4; Length 282;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	Db	2 GTTACTACTGAGC 15	44 GTTACTACTGAGC 57

```

RESULT 15
US-08-758-417A-413
Sequence 413, Application US/08/758417A
Patent No. 6300129
GENERAL INFORMATION:
APPLICANT: Lonberg, Nils
Kay, Robert M.
TITLE OF INVENTION: Transgenic No. 6300129-Human Animals for
Producing Heterologous Antibodies
NUMBER OF SEQUENCES: 417
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/758,417A
FILING DATE: 02-Dec-1996
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/728,463
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US 08/544,404
FILING DATE: 10-OCT-1995
APPLICATION NUMBER: US 08/352,322
FILING DATE: 07-DEC-1994
APPLICATION NUMBER: US 08/209,741
FILING DATE: 09-MAR-1994
APPLICATION NUMBER: US 08/165,699
FILING DATE: 10-DEC-1993
APPLICATION NUMBER: US 08/161,739
FILING DATE: 03-DEC-1993
APPLICATION NUMBER: US 08/155,301
FILING DATE: 18-NOV-1993
APPLICATION NUMBER: US 08/096,762
FILING DATE: 22-JUL-1993
APPLICATION NUMBER: US 08/053,131
FILING DATE: 26-APR-1993
APPLICATION NUMBER: US 07/990,860
FILING DATE: 16-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Serafini, Andrew T.
REGISTRATION NUMBER: 41,303
REFERENCE/DOCKET NUMBER: 014643-009030US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 413:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 413:
US-08-758-417A-413

```

Query Match 93.3%; Score 14; DB 4; Length 282;

Best Local Similarity 100.0%; Pred. No. 25;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	2	GTTACTACTGAGC	15
Db	44	GTTACTACTGAGC	57

Search completed: June 26, 2003, 04:52:44
Job time : 25.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:04:25 ; Search time 52.2 Seconds
(Without alignments)
421.674 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agttactactgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
8					
C 1	15	100.0	171	10	US-09-864-761-22522, A Sequence 22522, A
2	15	100.0	281	10	US-09-864-761-22248, A Sequence 22248, A
3	15	100.0	319	10	US-09-864-761-28401, A Sequence 28401, A
4	15	100.0	378	9	US-10-067-800-67, A Sequence 67, Appl
5	15	100.0	413	9	US-09-918-995-16699, A Sequence 16699, A
6	15	100.0	438	9	US-09-918-995-16650, A Sequence 16650, A
7	15	100.0	462	10	US-09-864-761-5474, A Sequence 5474, Ap
C 8	15	100.0	470	10	US-09-864-761-5760, A Sequence 5760, Ap
9	15	100.0	475	10	US-09-864-761-11831, A Sequence 11831, A
C 10	15	100.0	21470	9	US-10-092-154-1157, A Sequence 1157, Ap
C 11	15	100.0	21470	10	US-09-764-847-1157, A Sequence 1157, Ap
12	14	93.3	324	10	US-09-864-761-31244, A Sequence 31244, A
13	14	93.3	341	9	US-09-879-813-15, A Sequence 15, Appl
14	14	93.3	341	9	US-10-146-505-15, A Sequence 15, Appl
15	14	93.3	353	10	US-09-864-761-28159, A Sequence 28159, A
C 16	14	93.3	398	10	US-09-864-761-10457, A Sequence 10457, A
17	14	93.3	417	10	US-09-905-243-7, A Sequence 7, Appli
18	14	93.3	470	10	US-09-864-761-3704, A Sequence 3704, Ap
19	14	93.3	471	9	US-09-187-693-31, A Sequence 31, Appl

20	14	93.3	481	10	US-09-864-761-5338, A Sequence 5338, Ap
21	14	93.3	481	10	US-09-864-761-11572, A Sequence 11572, A
22	14	93.3	482	10	US-09-864-761-14715, A Sequence 14715, A
23	14	93.3	505	10	US-09-954-456-1183, A Sequence 1183, Ap
24	14	93.3	669	9	US-09-972-656-65, A Sequence 65, Appl
25	14	93.3	687	9	US-09-972-656-81, A Sequence 81, Appl
26	14	93.3	690	9	US-09-972-656-71, A Sequence 71, Appl
C 27	14	93.3	2624	9	US-09-854-133-19, A Sequence 19, Appl
C 28	14	93.3	2624	10	US-09-738-973-19, A Sequence 19, Appl
29	13.4	89.3	303	10	US-09-783-590-7401, A Sequence 7401, Ap
30	13.4	89.3	320	9	US-09-764-891-9777, A Sequence 9777, Ap
31	13.4	89.3	342	10	US-09-828-708-121, A Sequence 121, App
32	13.4	89.3	362	10	US-09-864-761-32297, A Sequence 32297, A
33	13.4	89.3	362	10	US-09-878-574-1765, A Sequence 1765, Ap
34	13.4	89.3	363	9	US-10-067-800-59, A Sequence 59, Appl
35	13.4	89.3	456	9	US-09-187-693-32, A Sequence 32, Appl
C 36	13.4	89.3	466	9	US-10-198-846-7483, A Sequence 7483, Ap
37	13.4	89.3	469	10	US-09-864-761-15792, A Sequence 15792, A
38	13.4	89.3	472	9	US-10-040-739-963, A Sequence 963, App
C 39	13.4	89.3	474	10	US-09-867-701-9902, A Sequence 9902, Ap
40	13.4	89.3	510	9	US-10-198-846-3994, A Sequence 3994, Ap
C 41	13.4	89.3	593	9	US-09-764-868-192, A Sequence 192, App
C 42	13.4	89.3	831	10	US-09-827-663-3, A Sequence 3, Appli
C 43	13.4	89.3	1278	10	US-09-815-242-9165, A Sequence 9165, Ap
C 44	13.4	89.3	1290	10	US-09-815-242-9491, A Sequence 9491, Ap
C 45	13.4	89.3	1293	10	US-09-827-663-1, A Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-09-864-761-22522/C

; Sequence 22522, Application US/09864761

; Patent No. US20020048763A1

; GENERAL INFORMATION:

; APPLICANT: Penn, Sharon G.

; APPLICANT: Rank, David R.

; APPLICANT: Hanzel, David K.

; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO

; FILE REFERENCE: Aecomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864, 761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180, 312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207, 456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632, 366

; PRIOR FILING DATE: 2000-08-03

; PRIOR APPLICATION NUMBER: GB 24263.6

; PRIOR FILING DATE: 2000-10-04

; PRIOR APPLICATION NUMBER: US 60/236, 359

; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00667

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00664

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00665

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00668

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00661

; PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22522
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004593.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 28
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
; OTHER INFORMATION: NT HIT: g18567391, EVALUOE 1.60e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE727658.1, EVALUOE 1.50e-01
; OTHER INFORMATION: SWISSPROT HIT: P07131, EVALUOE 7.80e-01
US-09-864-761-22522
```

```
Query Match      100.0%; Score 15; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AGTTACTACTGAGC 15
        |||||
Db      82 AGTTACTACTGAGC 68
```

```
RESULT 2
US-09-864-761-22248
; Sequence 22248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 22248
; LENGTH: 281
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB019441.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.9
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.3
; OTHER INFORMATION: EST_HUMAN HIT: AW405472.1, EVALUOE 0.00e+00
; OTHER INFORMATION: NT HIT: Z14240.1, EVALUOE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P06331, EVALUOE 2.00e-30
US-09-864-761-22248
```

```
Query Match      100.0%; Score 15; DB 10; Length 281;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 AGTTACTACTGAGC 15
        |||||
Db      64 AGTTACTACTGAGC 78
```

```
RESULT 3
US-09-864-761-28401
; Sequence 28401, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28401
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB019438.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56
; OTHER INFORMATION: NT HIT: U03896.1, EVALU0 0.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW402332.1, EVALU0 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P06331, EVALU0 3.00e-36
US-09-864-761-28401

Query Match          100.0%; Score 15; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||||||
Db      102 AGTTACTACTGAGC 116

RESULT 4
US-10-067-800-67
; Sequence 67, Application US/10067800
; Publication No. US20030100058A1
; GENERAL INFORMATION:
; APPLICANT: Roschke, Viktor
; APPLICANT: Rosen, Craig A.
; TITLE OF INVENTION: Human G-protein Chemokine Receptor (CCR5) HDGMR10
; FILE REFERENCE: 1488.1150001
; CURRENT APPLICATION NUMBER: US/10/067,800
; PRIOR APPLICATION NUMBER: PCT/US01/04153
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 09/779,880
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/297,257
; PRIOR FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/310,458
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/328,447
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/341,725
; PRIOR FILING DATE: 2001-12-21
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 378
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(378)
; OTHER INFORMATION:
US-10-067-800-67

Query Match          100.0%; Score 15; DB 9; Length 378;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||||||
Db      91 AGTTACTACTGAGC 105

RESULT 5
US-09-918-995-16699
; Sequence 16699, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16699
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(413)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-16699

Query Match          100.0%; Score 15; DB 9; Length 413;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||||||
Db      171 AGTTACTACTGAGC 185

RESULT 6
US-09-918-995-16650
; Sequence 16650, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16650
; LENGTH: 438
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-16650

Query Match          100.0%; Score 15; DB 9; Length 438;
```

Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
Db 166 AGTTACTACTGAGC 180

RESULT 7
US-09-864-761-5474

```
; Sequence 5474, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5474
; LENGTH: 462
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB019441.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 7.6
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 5.9
```

OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.3
US-09-864-761-5474

Query Match 100.0%; Score 15; DB 10; Length 462;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
Db 316 AGTTACTACTGAGC 330

RESULT 8
US-09-864-761-5760/c

```
; Sequence 5760, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR APPLICATION NUMBER: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 5760
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004593.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 28
```

```
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
US-09-864-761-5760
```

```
Query Match          100.0%; Score 15; DB 10; Length 470;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AGTTACTACTGAGC 15
          |||
Db      358 AGTTACTACTGAGC 344
```

```
RESULT 9
US-09-864-761-11831
; Sequence 11831, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11831
; LENGTH: 475
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; OTHER INFORMATION: MAP TO AB019438.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56
US-09-864-761-11831
```

```
Query Match          100.0%; Score 15; DB 10; Length 475;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AGTTACTACTGAGC 15
          |||
Db      329 AGTTACTACTGAGC 343
```

```
RESULT 10
US-10-092-154-1157/c
; Sequence 1157, Application US/10092154
; Publication No. US20030054375A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1157
; LENGTH: 21470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-092-154-1157
```

```
Query Match          100.0%; Score 15; DB 9; Length 21470;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AGTTACTACTGAGC 15
          |||
Db      10965 AGTTACTACTGAGC 10951
```

```
RESULT 11
US-09-764-847-1157/c
; Sequence 1157, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1157
; LENGTH: 21470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-847-1157
```

```
Query Match          100.0%; Score 15; DB 10; Length 21470;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 AGTTACTACTGAGC 15
          |||
Db      10965 AGTTACTACTGAGC 10951
```

```
RESULT 12
US-09-864-761-31244
; Sequence 31244, Application US/09864761
; Patent No. US20020048763A1
```



```

; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 31244
; LENGTH: 324
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB019439.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.44
; OTHER INFORMATION: NT HIT: M99683.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P06331, EVALUE 1.00e-26
; OTHER INFORMATION: EST_HUMAN HIT: AW403420.1, EVALUE 0.00e+00
US-09-864-761-31244

Query Match          93.3%; Score 14; DB 10; Length 324;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

; GENERAL INFORMATION:
; APPLICANT: Sale, Julian E.
; APPLICANT: Neuberger, Michael S.
; APPLICANT: Cumbers, Sarah J.
; TITLE OF INVENTION: Method of Generating Diversity
; FILE REFERENCE: 18396/2005
; CURRENT APPLICATION NUMBER: US/09/879,813
; CURRENT FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: 09/828,717
; PRIOR FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/GB99/03358
; PRIOR FILING DATE: 1999-10-08
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(341)
; OTHER INFORMATION:
US-09-879-813-15

```

```

Query Match          93.3%; Score 14; DB 9; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GTTACTACTGGAGC 15
        |||
Db       74 GTTACTACTGGAGC 87

```

```

RESULT 14
US-10-146-505-15
; Sequence 15, Application US/10146505
; Publication No. US20030108889A1
; GENERAL INFORMATION:
; APPLICANT: Sale, Julian E.
; APPLICANT: Neuberger, Michael S.
; APPLICANT: Cumbers, Sarah J.
; TITLE OF INVENTION: Method of Generating Diversity
; FILE REFERENCE: 18396/2005B
; CURRENT APPLICATION NUMBER: US/10/146,505
; CURRENT FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: 09/828,717
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 09/879,813
; PRIOR FILING DATE: 2001-06-11
; PRIOR APPLICATION NUMBER: PCT/GB99/03358
; PRIOR FILING DATE: 1999-10-08
; PRIOR APPLICATION NUMBER: GB 9822104.7
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: GB 9901141.3
; PRIOR FILING DATE: 1999-01-19
; PRIOR APPLICATION NUMBER: GB 9913435.5
; PRIOR FILING DATE: 1999-06-09
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(341)
; OTHER INFORMATION:
US-10-146-505-15

```

```

Query Match          93.3%; Score 14; DB 9; Length 341;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 2 GTTACTACTGAGC 15
| | | | | | | | | | | | | | | |
Db 74 GTTACTACTGAGC 87

RESULT 15
US-09-864-761-28159

; Sequence 28159, Application us/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28159
; LENGTH: 353
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AB019439.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.54
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.57
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.56
; OTHER INFORMATION: NT HIT: X92278.1, EVALUE 0.00e+00
; OTHER INFORMATION: SWISSPROT HIT: P06331, EVALUE 1.00e-53
; OTHER INFORMATION: EST_HUMAN HIT: BE672445.1, EVALUE 0.00e+00
US-09-864-761-28159

Query Match 93.3%; Score 14; DB 10; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 GTTACTACTGAGC 15
| | | | | | | | | | | | | | | |
Db 103 GTTACTACTGAGC 116

Search completed: June 26, 2003, 04:57:35
Job time : 61.2 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 26, 2003, 03:02:40 ; Search time 749.2 Seconds
(without alignments)
324.256 Million cell updates/sec

Title: US-09-138-091-42
Perfect score: 15
Sequence: 1 agtactactgagc 15

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	15	100.0	187	10	AW401419 UI-HF-BK0
2	15	100.0	248	10	AW403144 UI-HF-BK0
3	15	100.0	273	12	BF847860 IL5-EN008
4	15	100.0	275	10	AW408669 UI-HF-BM0
5	15	100.0	299	14	T27727 EST13874 Hu
6	15	100.0	319	10	AW405472 UI-HF-BL0

7	15	100.0	367	10	AW403544	AW403544	UI-HF-BK0
8	15	100.0	421	10	AW407630	AW407630	UI-HF-BM0
9	15	100.0	459	10	AW402533	AW402533	UI-HF-BK0
10	15	100.0	465	10	AW402332	AW402332	UI-HF-BK0
11	15	100.0	496	10	AW402282	AW402282	UI-HF-BK0
12	15	100.0	539	10	AW786572	AW786572	119835 MA
13	15	100.0	548	10	AW402602	AW402602	UI-HF-BK0
14	15	100.0	590	10	AV983832	AV983832	AV983832
15	15	100.0	607	14	BM783015	BM783015	K-EST0060
16	15	100.0	649	12	BG757507	BG757507	602714665
17	15	100.0	661	12	BG686421	BG686421	602638356
18	15	100.0	680	12	BG684306	BG684306	602636046
19	15	100.0	729	13	BI488497	BI488497	603021006
20	15	100.0	733	13	BM008087	BM008087	603617861
21	15	100.0	736	12	BG685592	BG685592	602637569
22	15	100.0	790	12	BG754804	BG754804	602714380
23	15	100.0	791	12	BG685529	BG685529	602637396
24	15	100.0	820	13	BI770055	BI770055	603053578
25	15	100.0	844	13	BI489640	BI489640	603032108
26	15	100.0	850	12	BG686474	BG686474	602637417
27	15	100.0	911	14	BQ711708	BQ711708	AGENCOURT
28	15	100.0	918	14	BQ708269	BQ708269	AGENCOURT
29	15	100.0	920	14	BQ710745	BQ710745	AGENCOURT
30	15	100.0	932	14	BQ710283	BQ710283	AGENCOURT
31	15	100.0	939	9	AL541900	AL541900	AL541900
32	15	100.0	947	14	BQ709527	BQ709527	AGENCOURT
33	15	100.0	951	14	BQ709739	BQ709739	AGENCOURT
34	15	100.0	964	12	BG757625	BG757625	602711208
35	15	100.0	971	9	AL560682	AL560682	AL560682
36	15	100.0	977	12	BG397634	BG397634	602438679
37	15	100.0	980	12	BG684861	BG684861	602636587
38	15	100.0	1005	14	BQ709036	BQ709036	AGENCOURT
39	15	100.0	1036	14	BQ072430	BQ072430	AGENCOURT
40	15	100.0	1108	14	BM920469	BM920469	AGENCOURT
41	15	100.0	1118	14	BQ882037	BQ882037	AGENCOURT
42	14	93.3	113	17	AZ013956	AZ013956	RPCT-23-3
43	14	93.3	186	14	BQ354864	BQ354864	MR1-HT106
44	14	93.3	196	10	AW401400	AW401400	UI-HF-BK0
45	14	93.3	214	13	BI225563	BI225563	602948847

ALIGNMENTS

RESULT 1
AW401419 187 bp mRNA linear EST 16-FEB-2000
LOCUS UI-HF-BK0-aaa-e-11-0-UI.r1 NIH-MGC_36 Homo sapiens cDNA clone
DEFINITION IMAGE:3053276 5', mRNA sequence.
ACCESSION AW401419
VERSION AW401419
KEYWORDS
SOURCE
ORGANISM
Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished (1999)
National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers
1. .187

FEATURES
Source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3053276"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      34 a      54 c      61 g      38 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 8.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||
Db      122 AGTTACTACTGAGC 136

RESULT 2
AW403144      248 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UI-HF-BK0-aar-g-10-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3054907 5', mRNA sequence.
ACCESSION      AW403144
VERSION      AW403144.1 GI:6921997
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 248)
TITLE      NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL      National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT      Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
FEATURES
source
1. 248
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3054907"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      55 a      73 c      69 g      51 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 248;
Best Local Similarity 100.0%; Pred. No. 9.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AGTTACTACTGAGC 15
        |||
Db      115 AGTTACTACTGAGC 129

RESULT 3
BF847860/c      273 bp      mRNA      linear      EST 16-JAN-2001
LOCUS      IL5-EN0085-021100-228-d07 EN0085 Homo sapiens cDNA, mRNA sequence.
DEFINITION      BF847860
ACCESSION      BF847860
VERSION      BF847860.1 GI:12235023
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 273)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&tl2=IL5-EN0085-
021100-228-d07&ts=2000-11-02&ta=1)
Seq primer: puc 18 forward
High quality sequence stop: 272.
FEATURES
source
1. 273
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="EN0085"
/dev_stage="Adult"
/note="Organ: lung_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      61 a      76 c      81 g      55 t
ORIGIN

Query Match      100.0%; Score 15; DB 12; Length 273;
Best Local Similarity 100.0%; Pred. No. 9.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGAGC 15
        |||
Db      175 AGTTACTACTGAGC 161

RESULT 4
AW408669      275 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UI-HF-BMO-ads-b-08-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
DEFINITION      IMAGE:3062726 5', mRNA sequence.
ACCESSION      AW408669

```


VERSION AW408669.1 GI:6927726
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 275)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers
1. 275
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3062726"
/clone_lib="NIH_MGC_38"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LT1)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(2.5-3.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 62 a 84 c 73 g 56 t
ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 275;
Best Local Similarity 100.0%; Pred. No. 9.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGAGC 15
|||||
Db 105 AGTTACTACTGAGC 119

RESULT 5
T27727 299 bp mRNA linear EST 06-SEP-1995
LOCUS EST13874 Human Testis Homo sapiens CDNA 5' end similar to
DEFINITION immunoglobulin mu (gamma) heavy chain, V(IV)DJC regions (HT:3057),
mRNA sequence.
T27727
T27727.1 GI:609825
EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 299)
AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult
C.J., Lee, N., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White
O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W., Clayton, R.A.,
Cline, R.T., Colton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald,
L.M., FitzHugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodek, A.,
Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle Jr, P.S., Kelley, J.M.,
Klimek, K.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,
Moreno-Palanges, R.F., McDonald, L.A., Nguyen, D.T., Pellegrino, S.M.,
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.-J.,

TITLE Dimke, D., Peng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.-W.,
Hu, J.-S., Greene, J.M., Gruber, J., Hudson, P., Kim, A., Kozak, D.L.,
Kunsch, C., Li, H., Li, H., Meissner, P.S., Olsen, H., Raymond, L., Wei
Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillon, P.J., Fannon
M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and
Venter, J.C.
JOURNAL Initial Assessment of Human Gene Diversity and Expression Patterns
MEDLINE Based Upon 83 Million Basepairs of CDNA Sequence
COMMENT Nature 377, 3-174 (1995)
96026280
Contact: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: tdbinfo@tdb.tigr.org
For clone availability, additional sequence and expression
information related to this EST, please contact the TIGR Database
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.
Location/Qualifiers
1. 299
/organism="Homo sapiens"
/db_xref="ATCC (inhost):108071"
/db_xref="taxon:9606"
/clone_lib="Human Testis"
/note="Organ: testis"

BASE COUNT 61 a 84 c 82 g 68 t 4 others
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 299;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGAGC 15
|||||
Db 169 AGTTACTACTGAGC 183

RESULT 6
AW405472 319 bp mRNA linear EST 16-FEB-2000
LOCUS UT-HF-BL0-adh-d-02-0-UT.r1 NIH_MGC_37 Homo sapiens CDNA clone
DEFINITION IMAGE:3061658 5', mRNA sequence.
AW405472
AW405472
AW405472.1 GI:6924529
EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 319)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.
Location/Qualifiers
1. 319
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3061658"
/clone_lib="NIH_MGC_37"
/tissue_type="lymph"

```
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(1.5-2.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      74 a      92 c      93 g      60 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 319;
Best Local Similarity 100.0%; Pred. No. 1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGGAGC 15
      |||
Db      100 AGTTACTACTGGAGC 114

RESULT 7
AW403544      367 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UI-HF-BK0-abf-d-09-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
ACCESSION      AW403544
VERSION      AW403544.1 GI:6922517
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE      1 (bases 1 to 367)
JOURNAL      NIH-MGC http://mgc.nci.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
source
Location/Qualifiers
1..367
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3055937"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      80 a      104 c      109 g      73 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 367;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGGAGC 15
      |||
Db      115 AGTTACTACTGGAGC 129
```

```
RESULT 8
AW407630      421 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UI-HF-BM0-adg-a-10-0-UI.r1 NIH_MGC_38 Homo sapiens cDNA clone
DEFINITION      IMAGE:3062322 5', mRNA sequence.
ACCESSION      AW407630
VERSION      AW407630.1 GI:6926697
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE      1 (bases 1 to 421)
JOURNAL      NIH-MGC http://mgc.nci.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
source
Location/Qualifiers
1..421
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3062322"
/clone_lib="NIH_MGC_38"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(2.5-3.5kb). Directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
BASE COUNT      91 a      131 c      111 g      87 t
ORIGIN

Query Match      100.0%; Score 15; DB 10; Length 421;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGTTACTACTGGAGC 15
      |||
Db      93 AGTTACTACTGGAGC 107

RESULT 9
AW402533      459 bp      mRNA      linear      EST 16-FEB-2000
LOCUS      UI-HF-BK0-aas-f-08-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
DEFINITION      IMAGE:3054878 5', mRNA sequence.
ACCESSION      AW402533
VERSION      AW402533.1 GI:6921226
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE      1 (bases 1 to 459)
JOURNAL      NIH-MGC http://mgc.nci.nih.gov/.
COMMENT      National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
```

Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
source

1. 459
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3054878"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 95 a 148 c 124 g 92 t
ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 459;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
|||||
DB 87 AGTTACTACTGAGC 101

RESULT 10 465 bp mRNA linear EST 16-FEB-2000
AM402332
LOCUS
DEFINITION
UT-HF-BK0-aal-b-06-0-UI.r1 NIH_MGC_36 Homo sapiens CDNA clone
IMAGE:3053915 5', mRNA sequence.

ACCESSION AM402332
VERSION
KEYWORDS
SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 465)
NIH-MGC <http://mgc.nci.nih.gov/>.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
source

1. 465
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3053915"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LTI)"

/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 102 a 137 c 146 g 80 t
ORIGIN

QY 1 AGTTACTACTGAGC 15
|||||
DB 122 AGTTACTACTGAGC 136

RESULT 11 496 bp mRNA linear EST 16-FEB-2000
AM402282
LOCUS
DEFINITION
UT-HF-BK0-aaj-e-06-0-UI.r1 NIH_MGC_36 Homo sapiens CDNA clone
IMAGE:3054035 5', mRNA sequence.

ACCESSION AM402282
VERSION
KEYWORDS
SOURCE
ORGANISM

human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 496)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
source

1. 496
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3054035"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/lab_host="DH10B (LTI)"
/note="Vector: pT7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 102 a 160 c 123 g 111 t
ORIGIN

Query Match 100.0%; Score 15; DB 10; Length 496;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGTTACTACTGAGC 15
|||||
DB 53 AGTTACTACTGAGC 67

RESULT 12
AW786572/c

LOCUS AW786572 539 bp mRNA linear EST 09-JUL-2000
 DEFINITION 119835 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
 ACCESSION AW786572
 VERSION AW786572.1 GI:7843348
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 1 (bases 1 to 539)
 Fahrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.
 Design and use of two pooled tissue normalized cDNA libraries for EST discovery in swine
 Unpublished (2000)
 CONTACT: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options.
 PCR Primers
 FORWARD: AGGAACACGTATGACCAT
 BACKWARD: GTTTCACAGTCACGACG
 Plate: 42 row: D column: 18
 Seq primer: ATTTAGTGACACTATAG.
 Location/Qualifiers
 1..539
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /clone_lib="MARC 1PIG"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI; Library made from pooled tissue from day 11, 13, 15, 20, and 30 embryos."
 BASE COUNT 164 a 101 c 142 g 132 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 539;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTTACTACTGAGC 15
 ||||||||||||||||
 Db 348 AGTTACTACTGAGC 334
 RESULT 13
 AW402602 548 bp mRNA linear EST 16-FEB-2000
 LOCUS UI-HF-BK0-aax-f-06-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
 DEFINITION IMAGE:3055259 5', mRNA sequence.
 ACCESSION AW402602
 VERSION AW402602.1 GI:6921304
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 548)
 NIH-MGC http://mgc.nci.nih.gov/.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs-remail.nih.gov
 Eco RI site shown at the beginning of the sequence.
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: M.B. Soares Lab

CDNA Library Arrayed by: M.B. Soares Lab
 DNA Sequencing by: M.B. Soares Lab
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/BLNL at: www-bio.lnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..548
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:3055259"
 /clone_lib="NIH_MGC_36"
 /tissue_type="lymph"
 /cell_type="germinal center B cells"
 /cell_line="MGC85"
 /lab_host="DH10B (LTI)"
 /note="Vector: p7T3-Pac; Site_1: NotI; Site_2: Eco RI; Constructed from size fractionated cytoplasmic mRNA (0.5-1.5kb). Directionally cloned. Cells provided by Louis M. Staudt, Ph.D. Library preparation by Maria de Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D."
 BASE COUNT 115 a 171 c 148 g 114 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 548;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTTACTACTGAGC 15
 ||||||||||||||||
 Db 109 AGTTACTACTGAGC 123
 RESULT 14
 AV983832 590 bp mRNA linear EST 14-MAR-2002
 LOCUS AV983832/c
 DEFINITION AV983832 Nori Satoh unpublished cDNA library, egg cDNA
 ACCESSION AV983832
 VERSION AV983832.1 GI:19472913
 KEYWORDS EST.
 SOURCE Ciona intestinalis.
 ORGANISM Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Clonidae; Ciona.
 1 (bases 1 to 590)
 Satoh, N., Satou, Y., Kohara, Y. and Shin-i, T.
 Expressed genes in Ciona intestinalis
 Unpublished (2000)
 CONTACT: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-ku, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 Location/Qualifiers
 1..590
 /organism="Ciona intestinalis"
 /db_xref="taxon:7719"
 /clone="cieg28e23"
 /clone_lib="Nori Satoh unpublished cDNA library, egg"
 /tissue_type="whole animal"
 /dev_stage="egg"
 /note="Vector: pBluescript SK"
 BASE COUNT 200 a 152 c 101 g 137 t
 ORIGIN
 Query Match 100.0%; Score 15; DB 10; Length 590;
 Best Local Similarity 100.0%; Pred. No. 1.3e+03;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AGTTACTACTGAGC 15

DB 243 AGTTACTACTGAGC 229

RESULT 15
BM783015 607 bp mRNA linear EST 05-MAR-2002
LOCUS K-EST0060735 S18N669761 Homo sapiens cDNA clone S18N669761-2-E06
DEFINITION 5', mRNA sequence.
ACCESSION BM783015
VERSION BM783015.1 GI:19131247
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 607)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 2 row: E column: 06
High quality sequence stop: 607.

FEATURES
source
1. 607
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S18N669761-2-E06"
/clone_lib="S18N669761"
/sex="F"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dT-selected mRNA by
priming with dT-tailed vector. The dT-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 131 a 187 c 163 g 126 t
ORIGIN

Query Match 100.0%; Score 15; DB 14; Length 607;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGTTACTACTGAGC 15
DB 193 AGTTACTACTGAGC 207

Search completed: June 26, 2003, 04:50:22
Job time : 753.2 secs

